- Higher eukaryotes transcription initiation factor TFIID 250 Kd subunit (TBP-associated factor p250) (gene CCG1). P250 associated with the TFIID TATA-box binding protein and seems essential for progression of the G1 p: hase of the cell cycle.
- Human RING3, a protein of unknown funct. ion encoded in the MHC class II locus.
- Mammalian CREB-binding protein (CBP), wh. ich mediates cAMP-gene regulation by binding specifically to phosphorylated CREB protein.
- Drosophila female sterile homeotic protein (gene fsh). Evequired maternally for proper expression of other homeotic genes involved in pattern formanteen, such as Ubx.
- Drosophila brahma protein (gene brm), a protein required for t. he activation of multiple homeotic genes.
 - Mammalian homologs of brahma. In human, three brahma-like proteins are known: SNF2a(hBRM), SNF2b, and BRG1.
 - Human BS69, a protein that binds to adenovirus E1A and inhibits E1A transactivation
- Human peregrin (or Br140).

- Yeast BDF1 [3], a transcription factor involved in the expression of a broad class of genes including snRNAs.
- Yeast GCN5, a general transcriptional activator operating in concert with certain other DNA-binding transcriptional activators, such as GCN4, HAP2/3/4 or ADA2.
- 20 Yeast NPS1/STH1, involved in G(2) phase control in mitosis.
 - Yeast SNF2/SWI2, which is part of a complex with the SNF5, SNF6, SWI3 and ADR6/SWI1 proteins. This SWI-complex is involved in transcriptional activation.
 - Yeast SPT7, a transcriptional activator of Ty elements and possibly other genes.
 - Caenorhabditis elegans protein cbp-1.
- 25 Yeast hypothetical protein YGR056w.
 - Yeast hypothetical protein YKR008w.
 - Yeast hypothetical protein L9638.1.

Some proteins contain a region which, while similar to some extent to a classical bromodomain, diverges from it by either lacking part of the domain or because of an insertion. These proteins are:

15

5

- Mammalian protein HRX (also known as All-1 or MLL), a protein involved in translocations leading to acute leukemias and which possibly acts as a transcriptional
- regulatory factor. HRX contains a region similar to the C-terminal half of the bromodomain.
- Caenorhabditis elegans hypothetical protein ZK783.4. The bromodomain of this protein has a 23 amino-acid insertion.
- Yeast protein YTA7. This protein contains a region with significant similarity to the Cterminal half of the bromodomain. As it is a member of the AAA family (see <PDOC00572>) it is also in a functionally different context.
- 10 The above proteins generally contain a single bromodomain, but some o them contain two copies, this is the case of BDF1, CCG1, fsh, RING3. YKR008w and L9638.1.

The exact function of this domain is not yet known but it is thought to be involved in proteinprotein interactions and it may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation.

The consensus pattern that has been developed spans a major part of the bromodomain; a more sensitive detection is available through the use of a profile which spans the whole domain.

Consensus pattern[STANVF]-x(2)-F-x(4)-[DNS]-x(5,7)-[DENGTF]-Y-[HFY]-x(2)-[LIVMFY]-x(3)-[LIVM]-x(4)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(6,8)-X(12,13)-[LIVx(2)-N-[SACF]-x(2)-[FY]

25 References

- [1] Haynes S.R., Doolard C., Winston F., Beck S., Trowsdale J., Dawid I.B. Neucleic Acids Res. 20:2693-2603(1992).
- [2] Tamkun J.W., Deuring R., Scott M.P., Kissinger M., Pattatucci A.M., Kauft nan T.C., Kennison J.A. Cell 68:561-572(1992).
- [3] Tamkun J.W. Curr. Opin. Genet. Dev. 5:473-477(1959), 30

808. (CH) Actinin-type actin-binding domain signatures PROSITE cross-reference(s): PS00019; ACTININ_1, PS00020; ACTININ_2



Alpha-actinin is a F-actin cross-linking protein which is thought to anchoractin to a variety of intracellular structures [1]. The actin-binding domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]:

5

- In the beta chain of spectrin (or fodrin).
- In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane.
- In the slime mold gelation factor (or ABP-120).
- In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to membrane glycoproteins.
 - In fimbrin (or plastin), an actin-bundling protein. Fimbrin differs from the above proteins in that it contains two tandem copies of the actin-binding domain and that these copies are located in the C-terminal part of the protein.

15

Two conserved regions were selected as signature patterns for this type of main. The first of this region is located at the beginning of the domain, hile the second one is located in the central section and has been shown to be essential for the binding of actin.

20

- Consensus pattern[EQ]-x(2)-[ATV]-[FY]-x(2)-W-x-N

 Consensus pattern[LIVM]-x-[SGN]-[LIVM]-[DAGHE]-[SAG]-x-[DNEAG]-[LIVM]-x[DEAG]-x(4)-[LIVM]-x-[LM]-[SAG]-[LIVM]-[LIVMT]-W-x- [LIVM](2)
- [1] Schleicher M., Andre E., Harmann A., Noegel A.A. Dev. Genet. 9:521-530(1988).
- 25 [2] Matsudaira P. Trends Biochem. Sci. 16:87-92(1991).
 - [3] Dubreuil R.R. BioEssays 13:219-226(1991).
 - 809. (COX1) Heme-copper oxidase subunit I, copper B binding region signature PROSITE cross-reference(s): PS00077; COX1
- Heme-copper respiratory oxidases [1] are oligomeric integral membrane protein complexes that catalyze the terminal step in the respiratory chain: they transfer electrons from cytochrome c or a quinol to oxygen. Some terminal oxidases generate a transmembrane proton gradient across the plasma membrane (prokaryotes) or the mitochondrial inner membrane (eukaryotes). The enzyme

10

15

30

complex consists of 3-4 subunits (prokaryotes) up to 13 polypeptides (mammals) of which only the catalytic subunit (equivalent to mammalian subunit 1 (CO I)) is found in all heme-copper respiratory oxidases. The presence of a bimetallic center (formed by a high-spin heme and copper B) as well as a low-spin heme, both ligated to six conserved histidine residues near the outer side of four transmembrane spans within CO I is common to all family members [2-4].

In contrary to eukaryotes the respiratory chain of prokaryotes is branched to multiple terminal oxidases. The enzyme complexes vary in heme and copper composition, substrate type and substrate affinity. The different respiratory oxidases allow the cells to customize their respiratory systems according a variety of environmental growth conditions [1].

Recently also a component of an anaerobic respiratory chain has been found to contain the copper B binding signature of this family: nitric oxide reductase (NOR) exists in denitrifying species of Archae and Eubacteria.

Enzymes that belong to this family are:

- Mitochondrial-type cytochrome c oxidase (EC 1.9.3.1) which uses cytochrome c as electron donor. The electrons are transferred via copper A (Cu(A)) and heme a to the bimetallic center of CO I that is formed by a pentacoordinated heme a and copper B (Cu(B)). Subunit 1 contains 12 transmembrane regions. Cu(B) is said to be ligated to three of the
 conserved histidine residues within the transmembrane segments 6 and 7.
 - Quinol oxidase from prokaryotes that transfers electrons from a quinol to the binuclear center of polypeptide I. This category of enzymes includes Escherichia coli cytochrome O terminal oxidase complex which is a component of the aerobic respiratory chain that predominates when cells are grown at high aeration.
 - FixN, the catalytic subunit of a cytochrome c oxidase expressed in nitrogen-fixing bacteroids living in root nodules. The high affinity for oxygen allows oxidative phosphorylation under low oxygen concentrations. A similar enzyme has been found in other purple bacteria.

- Nitric oxide reductase (EC 1.7.99.7) from Pseudomonas stutzeri. NOR reduces nitrate to dinitrogen. It is a heterodimer of norC and the catalytic subunit norB. The latter contains the 6 invariant histidine residues and 12 transmembrane segments [5].

5

As a signature pattern the copper-binding region was used.

Consensus pattern[YWG]-[LIVFYWTA](2)-[VGS]-H-[LNP]-x-V-x(44,47)-H-H [The three H's are copper B ligands]

10

Notecytochrome bd complexes do not belong to this family.

[1]

Garcia-Horsman J.A., Barquera B., Rumbley J., Ma J., Gennis R.B.

15 J. Bacteriol. 176:5587-5600(1994).

[2]

Castresana J., Luebben M., Saraste M., Higgins D.G.

EMBO J. 13:2516-2525(1994).

[3]

20

Capaldi R.A., Malatesta F., Darley-Usmar V.M. Biochim. Biophys. Acta 726:135-148(1983).

[4]

Holm L., Saraste M., Wikstrom M.

EMBO J. 6:2819-2823(1987).

25 **[5]**

Saraste M., Castresana J.

FEBS Lett. 341:1-4(1994).

810. (dehydrog_molyb) Eukaryotic molybdopterin oxidoreductases signature

PROSITE cross-reference(s): PS00559; MOLYBDOPTERIN_EUK

A number of different eukaryotic oxidoreductases that require and bind a molybdopterin cofactor have been shown [1] to share a few regions of sequence similarity. These enzymes are:

10

15

- Xanthine dehydrogenase (EC 1.1.1.204), which catalyzes the oxidation of xanthine to uric acid with the concomitant reduction of NAD. Structurally, this enzyme of about 1300 amino acids consists of at least three distinct domains: an N-terminal 2Fe-2S ferredoxin-like iron-sulfur binding domain (see <PDOC00175>), a central FAD/NAD-binding domain and a C-terminal Mopterin domain.
- Aldehyde oxidase (EC 1.2.3.1), which catalyzes the oxidation aldehydes into acids. Aldehyde oxidase is highly similar to xanthine dehydrogenase in its sequence and domain structure.
- Nitrate reductase (EC 1.6.6.1), which catalyzes the reduction of nitrate to nitrite. Structurally, this enzyme of about 900 amino acids consists of an N-terminal Mo-pterin domain, a central cytochrome b5-type heme-binding domain (see <PDOC00170>) and a C-terminal FAD/NAD-binding cytochrome reductase domain.
- Sulfite oxidase (EC 1.8.3.1), which catalyzes the oxidation of sulfite to sulfate. Structurally, this enzyme of about 460 amino acids consists of an N-terminal cytochrome b5-binding domain followed by a Mo-pterin domain.
- There are a few conserved regions in the sequence of the molybdopterin-binding domain of these enzymes. The pattern uses to detect these proteins is based on one of them. It contains a cysteine residue which could be involved in binding the molybdopterin cofactor.
- Consensus pattern[GA]-x(3)-[KRNQHT]-x(11,14)-[LIVMFYWS]-x(8)-[LIVMF]-x-C-x(2)-[DEN]-x(2)-[DE]

[1]

30

Wootton J.C., Nicolson R.E., Cock J.M., Walters D.E., Burke J.F., Doyle W.A., Bray R.C.
Biochim. Biophys. Acta 1057:157-185(1991).

811. (DNA_ligase) ATP-dependent DNA ligase signatures
PROSITE cross-reference(s): PS00697; DNA LIGASE A1, PS00333; DNA LIGASE A2

10

15

20

25

DNA ligase (polydeoxyribonucleotide synthase) is the enzyme that joins two DNA fragments by catalyzing the formation of an internucleotide ester bond between phosphate and deoxyribose. It is active during DNA replication, DNA repair and DNA recombination. There are two forms of DNA ligase: one requires ATP (EC 6.5.1.1), the other NAD (EC 6.5.1.2).

Eukaryotic, archaebacterial, virus and phage DNA ligases are ATP-dependent. During the first step of the joining reaction, the ligase interacts with ATP to form a covalent enzyme-adenylate intermediate. A conserved lysine residue is the site of adenylation [1,2].

Apart from the active site region, the only conserved region common to all ATP-dependent DNA ligases is found [3] in the C-terminal section and contains a conserved glutamate as well as four positions with conserved basic residues.

Signature patterns were developed for both conserved regions.

Consensus pattern[EDQH]-x-K-x-[DN]-G-x-R-[GACIVM] [K is the active site residue]

Consensus patternE-G-[LIVMA]-[LIVM](2)-[KR]-x(5,8)-[YW]-[QNEK]-x(2,6)-[KRH]-x(3,5)-K-[LIVMFY]-K

Sequences known to belong to this class detected by the patternALL, except for archebacterial DNA ligases.

[1]

Tomkinson A.E., Totty N.F., Ginsburg M., Lindahl T.

Proc. Natl. Acad. Sci. U.S.A. 88:400-404(1991).

30 [2]

Lindahl T., Barnes D.E.

Annu. Rev. Biochem. 61:251-281(1992).

[3]

Kletzin A.

10

Nucleic Acids Res. 20:5389-5396(1992).

812. (FAD_Gly3P_dh) FAD-dependent glycerol-3-phosphate dehydrogenase signatures PROSITE cross-reference(s): PS00977; FAD_G3PDH_1, PS00978; FAD_G3PDH_2

FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) (GPD) catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate. In bacteria [1] it is associated with the utilization of glycerol coupled to respiration. In Escherichia coli, two isozymes are known: one expressed under anaerobic conditions (gene glpA) and one in aerobic conditions (gene glpD). In eukaryotes, a mitochondrial form of GPD participates in the glycerol phosphate shuttle in conjunction with an NAD-dependent cytoplasmic GPD (EC 1.1.1.8) [2, 3].

These enzymes are proteins of about 60 to 70 Kd which contain a probable FAD-binding domain in their N-terminal extremity. The mammalian enzyme differs from the bacterial or yeast proteins by having an EF-hand calcium-binding region (See <PDOC00018>) in its C-terminal extremity.

Two signature patterns were developed. One based on the first half of the FADbinding domain and one which corresponds to a conserved region in the central part of these enzymes.

Consensus pattern[IV]-G-G-G-x(2)-G-[STACV]-G-x-A-x-D-x(3)-R-G

Consensus patternG-G-K-x(2)-[GSTE]-Y-R-x(2)-A

[1]

Austin D., Larson T.J.

J. Bacteriol. 173:101-107(1991).

30 [2]

25

Roennow B., Kielland-Brandt M.C.

Yeast 9:1121-1130(1993).

[3]

Brown L.J., McDonald M.J., Lehn D.A., Moran S.M.

10

15

J. Biol. Chem. 269:14363-14366(1994).

813. (Fapy_DNA_glyco) Formamidopyrimidine-DNA glycosylase signature PROSITE cross-reference(s): PS01242; FPG

Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) [1] (Fapy-DNA glycosylase) (gene fpg) is a bacterial enzyme involved in DNA repair and which excise oxidized purine bases to release 2,6-diamino-4-hydroxy-5N-methylformamidopyrimidine (Fapy) and 7,8-dihydro-8-oxoguanine (8-OxoG) residues. In addition to its glycosylase activity, FPG can also nick DNA at apurinic/apyrimidinic sites (AP sites). FPG is a monomeric protein of about 32 Kd which binds and require zinc for its activity.

The binding site for zinc seems to be located in the C-terminal part of the enzyme where fours conserved and essential [2] cysteines are located. A signature pattern was developed based on this region.

Consensus patternC-x(2,4)-C-x-[GTAQ]-x-[IV]-x(7)-R-[GSTAN]-[STA]-x-[FYI]-C- x(2)-C-Q

[The four C's are putative zinc ligands]

[1]

Duwat P., de Oliveira R., Ehrlich S.D., Boiteux S.

Microbiology 141:411-417(1995).

25 **[2]**

30

O'Connor T.E., Graves R.J., Demurcia G., Castaing B., Laval J.

J. Biol. Chem. 268:9063-9070(1993).

814. (G_glu_transpept) Gamma-glutamyltranspeptidase signature PROSITE cross-reference(s): PS00462; G_GLU_TRANSPEPTIDASE

Gamma-glutamyltranspeptidase (EC 2.3.2.2) (GGT) [1] catalyzes the transfer of the gamma-glutamyl moiety of glutathione to an acceptor that may be an amino acid, a peptide or water (forming glutamate). GGT plays a key role in the

10

15

20

30

gamma-glutamyl cycle, a pathway for the synthesis and degradation of glutathione. In prokaryotes and eukaryotes, it is an enzyme that consists of two polypeptide chains, a heavy and a light subunit, processed from a single chain precursor. The active site of GGT is known to be located in the light subunit.

The sequences of mammalian and bacterial GGT show a number of regions of high similarity [2]. Pseudomonas cephalosporin acylases (EC 3.5.1.-) that convert 7-beta-(4-carboxybutanamido)-cephalosporanic acid (GL-7ACA) into 7-aminocephalosporanic acid (7ACA) and glutaric acid are evolutionary related to GGT and also show some GGT activity [3]. Like GGT, these GL-7ACA acylases, are also composed of two subunits.

One of the conserved regions correspond to the N-terminal extremity of the mature light chains of these enzymes. This region was used as a signature pattern.

Consensus patternT-[STA]-H-x-[ST]-[LIVMA]-x(4)-G-[SN]-x-V-[STA]-x-T-x-T-[LIVM]-[NE]-x(1,2)-[FY]-G

[1]

Tate S.S., Meister A.

Meth. Enzymol. 113:400-419(1985).

[2]

Suzuki H., Kumagai H., Echigo T., Tochikura T.

J. Bacteriol. 171:5169-5172(1989).

[3]

Ishiye M., Niwa M.

Biochim. Biophys. Acta 1132:233-239(1992).

815. G-protein gamma subunit profile
PROSITE cross-reference(s): PS50058; G_PROTEIN_GAMMA

Guanine nucleotide-binding proteins (G proteins) [1] act as intermediaries in

10

the transduction of signals generated by transmembrane receptors. G proteins consist of three subunits (alpha, beta, and gamma). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

The gamma subunits are small proteins (from 70 to 110 residues) that are bound to the membrane via a isoprenyl group (either a farnesyl or a geranylgeranyl) covalently linked to their C-terminus. In mammals there are at least 12 different isoforms of gamma subunits.

The Caenorhabditis elegans protein egl-10, which is a regulator of G-protein signalling, contains a G-protein gamma-like domain.

A profile was developed that spans the complete length of the gamma subunit.

[1]

30

Pennington S.R.

20 Protein Prof. 2:16-315(1995).

816. GNS1/SUR4 family signature PROSITE cross-reference(s): PS01188; GNS1_SUR4

- The following group of eukaryotic integral membrane proteins, whose exact function has not yet clearly been established, are evolutionary related [1]:
 - Yeast GNS1 [2], a protein involved in synthesis of 1,3-beta-glucan.
 - Yeast SUR4 (or APA1, SRE1) [3], a protein that could act in a glucosesignaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.
 - Yeast hypothetical protein YJL196c.
 - Caenorhabditis elegans hypothetical protein C40H1.4.
 - Caenorhabditis elegans hypothetical protein D2024.3.

The proteins have from 290 to 435 amino acid residues. Structurally, they seem to be formed of three sections: a N-terminal region with two transmembrane domains, a central hydrophilic loop and a C-terminal region that contains from one to three transmembrane domains. A conserved region that contains three histidines was selected as a signature pattern. This region is located in the hydrophilic loop.

Consensus patternL-x-F-L-H-x-Y-H-H

10

5

[1]

Bairoch A.

Unpublished observations (1996).

[2]

15 El-Sherbeini M., Clemas J.A.

J. Bacteriol. 177:3227-3234(1995).

[3]

Garcia-Arranz M., Maldonado A.M., Mazon M.J., Portillo F.

J. Biol. Chem. 269:18076-18082(1994).

20

817. Immunoglobulins and major histocompatibility complex proteins signature PROSITE cross-reference(s): PS00290; IG_MHC

The basic structure of immunoglobulin (Ig) [1] molecules is a tetramer of two light chains and two heavy chains linked by disulfide bonds. There are two types of light chains: kappa and lambda, each composed of a constant domain (CL) and a variable domain (VL). There are five types of heavy chains: alpha, delta, epsilon, gamma and mu, all consisting of a variable domain (VH) and three (in alpha, delta and gamma) or four (in epsilon and mu) constant domains (CH1 to CH4).

The major histocompatibility complex (MHC) molecules are made of two chains. In class I [2] the alpha chain is composed of three extracellular domains, a transmembrane region and a cytoplasmic tail. The beta chain (beta-2-

20

microglobulin) is composed of a single extracellular domain. In class II [3], both the alpha and the beta chains are composed of two extracellular domains, a transmembrane region and a cytoplasmic tail.

It is known [4,5] that the Ig constant chain domains and a single extracellular domain in each type of MHC chains are related. These homologous domains are approximately one hundred amino acids long and include a conserved intradomain disulfide bond. A small pattern around the C-terminal cysteine is involved in this disulfide bond which can be used to detect these category of Ig related proteins.

Consensus pattern[FY]-x-C-x-[VA]-x-H-Sequences known to belong to this class detected by the pattern: Ig heavy chains type Alpha C region: All, in CH2 and CH3. Ig heavy chains type Delta C region: All, in CH3. Ig heavy chains type Epsilon C region: All, in CH1, CH3 and CH4. Ig heavy chains type Gamma C region: All, in CH3 and also CH1 in some cases Ig heavy chains type Mu C region: All, in CH2, CH3 and CH4. Ig light chains type Kappa C region: In all CL except rabbit and Xenopus. Ig light chains type Lambda C region: In all CL except rabbit. MHC class I alpha chains: All, in alpha-3 domains, including in the cytomegalovirus MHC-1 homologous protein [6]. Beta-2-microglobulin: All. MHC class II alpha chains: All, in alpha-2 domains. MHC class II beta chains: All, in beta-2 domains.

[1]

25 Gough N.

Trends Biochem. Sci. 6:203-205(1981).

[2]

Klein J., Figueroa F.

Immunol. Today 7:41-44(1986).

30 [3]

Figueroa F., Klein J.

Immunol. Today 7:78-81(1986).

[4]

Orr H.T., Lancet D., Robb R.J., Lopez de Castro J.A., Strominger J.L.

[5]

Cushley W., Owen M.J.

Nature 282:266-270(1979).

Immunol. Today 4:88-92(1983).

5 [6]

10

15

Beck S., Barrel B.G.

Nature 331:269-272(1988).

818. (IGFBP) Insulin-like growth factor binding proteins signature PROSITE cross-reference(s): PS00222; IGF BINDING

The insulin-like growth factors (IGF-I and IGF-II) bind to specific binding proteins in extracellular fluids with high affinity [1,2,3]. These IGF-binding proteins (IGFBP) prolong the half-life of the IGFs and have been shown to either inhibit or stimulate the growth promoting effects of the IGFs on cells culture. They seem to alter the interaction of IGFs with their cell surface receptors. There are at least six different IGFBPs and they are structurally related.

- The following growth-factor inducible proteins are structurally related to IGFBPs and could function as growth-factor binding proteins [4,5]:
 - Mouse protein cyr61 and its probable chicken homolog, protein CEF-10.
 - Human connective tissue growth factor (CTGF) and its mouse homolog, protein FISP-12.
 - Vertebrate protein NOV.

As a signature pattern a conserved cysteine-rich region located in the N-terminal section of these proteins is used.

Consensus patternG-C-[GS]-C-C-x(2)-C-A-x(6)-C

Sequences known to belong to this class detected by the patternALL, except for IGFBP-6's.

30

25

[1]

Rechler M.M.

Vitam. Horm. 47:1-114(1993).

[2]

5 Shimasaki S., Ling N.

Prog. Growth Factor Res. 3:243-266(1991).

[3]

Clemmons D.R.

Trends Endocrinol. Metab. 1:412-417(1990).

10 [4]

Bradham D.M., Igarashi A., Potter R.L., Grotendorst G.R.

J. Cell Biol. 114:1285-1294(1991).

[5]

Maloisel V., Martinerie C., Dambrine G., Plassiart G., Brisac M., Crochet

15 J., Perbal B.

Mol. Cell. Biol. 12:10-21(1992).

819. LMWPc: Low molecular weight phosphotyrosine protein phosphatase

Number of members: 34

20

30

[1]Medline: 94329182, The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. Su XD, Taddei N, Stefani M, Ramponi G, Nordlund P; Nature 1994;370:575-578.

25 820. (myosin_head) ATP/GTP-binding site motif A (P-loop)
PROSITE cross-reference(s): PS00017; ATP_GTP_A

From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5].

There are numerous ATP- or GTP-binding proteins in which the P-loop is found. A number of protein families for which the relevance of the presence of such motif has been noted is listed below:

5

- ATP synthase alpha and beta subunits (see <PDOC00137>).
- Myosin heavy chains.
- Kinesin heavy chains and kinesin-like proteins (see <PDOC00343>).
- Dynamins and dynamin-like proteins (see <PDOC00362>).
- Guanylate kinase (see <PDOC00670>).
 - Thymidine kinase (see <PDOC00524>).
 - Thymidylate kinase (see <PDOC01034>).
 - Shikimate kinase (see <PDOC00868>).
 - Nitrogenase iron protein family (nifH/frxC) (see <PDOC00580>).
- ATP-binding proteins involved in 'active transport' (ABC transporters) [7] (see <PDOC00185>).
 - DNA and RNA helicases [8,9,10].
 - GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.).
 - Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.).
- Nuclear protein ran (see <PDOC00859>).
 - ADP-ribosylation factors family (see <PDOC00781>).
 - Bacterial dnaA protein (see <PDOC00771>).
 - Bacterial recA protein (see <PDOC00131>).
 - Bacterial recF protein (see <PDOC00539>).
- Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.).
 - DNA mismatch repair proteins mutS family (See <PDOC00388>).
 - Bacterial type II secretion system protein E (see <PDOC00567>).

Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved for

adenylate kinase, in which there is a single deviation from the P-loop pattern: in the last position Gly is found instead of Ser or Thr.

Consensus pattern[AG]-x(4)-G-K-[ST]

5

[1]

Walker J.E., Saraste M., Runswick M.J., Gay N.J.

EMBO J. 1:945-951(1982).

[2]

10 Moller W., Amons R.

FEBS Lett. 186:1-7(1985).

[3]

Fry D.C., Kuby S.A., Mildvan A.S.

Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986).

15 [4]

Dever T.E., Glynias M.J., Merrick W.C.

Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).

[5]

Saraste M., Sibbald P.R., Wittinghofer A.

20 Trends Biochem. Sci. 15:430-434(1990).

[6]

Koonin E.V.

J. Mol. Biol. 229:1165-1174(1993).

[7]

Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P.

J. Bioenerg. Biomembr. 22:571-592(1990).

[8]

Hodgman T.C.

30 Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).

[9]

Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K.,

Schnier J., Slonimski P.P.

Nature 337:121-122(1989).

15

20

30

[10]

Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989).

5 **821. PE: PE family**

This family named after a PE motif near to the amino terminus of the domain. The PE family of proteins all contain an amino-terminal region of about 110 amino acids. The carboxyl terminus of this family are variable and fall into several classes. The largest class of PE proteins is the highly repetitive PGRS class which have a high glycine content. The function of these proteins is uncertain but it has been suggested that they may be related to antigenic variation of Mycobacterium tuberculosis [1]. Number of members: 88

[1] Medline: 98295987. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG, et al; Nature 1998;393:537-544.

822. (RNB) Ribonuclease II family signature
PROSITE cross-reference(s): PS01175; RIBONUCLEASE_II

On the basis of sequence similarities, the following bacterial and eukaryotic proteins seem to form a family:

- Escherichia coli and related bacteria ribonuclease II (EC 3.1.13.1) (RNase II) (gene rnb) [1]. RNase II is an exonuclease involved in mRNA decay. It degrades mRNA by hydrolyzing single-stranded polyribonucleotides processively in the 3' to 5' direction.
 - Bacterial protein vacB. In Shigella flexneri, vacB has been shown to be required for the expression of virulence genes at the posttranscriptional level.
 - Yeast protein SSD1 (or SRK1) which is implicated in the control of the cell cycle G1 phase.
 - Yeast protein DIS3 [2], which binds to ran (GSP1) and ehances the the

669

nucleotide-releasing activity of RCC1 on ran.

- Fission yeast protein dis3, which is implicated in mitotic control.
- Neurospora crassa cyt-4, a mitochondrial protein required for RNA 5' and 3' end processing and splicing.
- 5 Yeast protein MSU1, which is involved in mitochondrial biogenesis.
 - Synechocystis strain PCC 6803 protein zam [3], which control resistance to the carbonic anhydrase inhibitor acetazolamide.
 - Caenorhabditis elegans hypothetical protein F48E8.6.
- The size of these proteins range from 644 residues (rnb) to 1250 (SSD1). While their sequence is highly divergent they share a conserved domain in their C-terminal section [4]. It is possible that this domain plays a role in a putative exonuclease function that would be common to all these proteins. A signature pattern was developed based on the core of this conserved domain.

Consensus pattern[HI]-[FYE]-[GSTAM]-[LIVM]-x(4,5)-Y-[STAL]-x-[FWVAC]-[TV]-[SA]-P-[LIVMA]-[RQ]-[KR]-[FY]-x-D-x(3)-[HQ]

[1]

Zilhao R., Camelo L., Arraiano C.M.Mol. Microbiol. 8:43-51(1993).

[2]

Noguchi E., Hayashi N., Azuma Y., Seki T., Nakamura M., Nakashima N.,

Yanagida M., He X., Mueller U., Sazer S., Nishimoto T.

25 EMBO J. 15:5595-5605(1996).

[3]

Beuf L., Bedu S., Cami B., Joset F.

Plant Mol. Biol. 27:779-788(1995).

[4]

30 Mian I.S.

Nucleic Acids Res. 25:3187-3195(1997).

823. Src homology 2 (SH2) domain profile

PROSITE cross-reference(s): PS50001; SH2

20

5

The Src homology 2 (SH2) domain is a protein domain of about 100 amino-acid residues first identified as a conserved sequence region between the oncoproteins Src and Fps [1]. Similar sequences were later found in many other intracellular signal-transducing proteins [2]. SH2 domains function as regulatory modules of intracellular signalling cascades by interacting with high affinity to phosphotyrosine-containing target peptides in a sequence-specific and strictly phosphorylation-dependent manner [3,4,5,6].

The SH2 domain has a conserved 3D structure consisting of two alpha helices and six to seven beta-strands. The core of the domain is formed by a continuous beta-meander composed of two connected beta-sheets [7].

So far, SH2 domains have been identified in the following proteins:

- Many vertebrate, invertebrate and retroviral cytoplasmic (non-receptor) protein tyrosine kinases. In particular in the Src, Abl, Bkt, Csk and ZAP70 families of kinases.
- Mammalian phosphatidylinositol-specific phospholipase C gamma-1 and -2. Two copies of the SH2 domain are found in those proteins in between the catalytic 'X-' and 'Y-boxes' (see <PDOC50007>).
- Mammalian phosphatidyl inositol 3-kinase regulatory p85 subunit.
- Some vertebrate and invertebrate protein-tyrosine phosphatases.
- Mammalian Ras GTPase-activating protein (GAP).
- Adaptor proteins mediating binding of guanine nucleotide exchange factors to growth factor receptors: vertebrate GRB2, Caenorhabditis elegans sem-5 and Drosophila DRK.
 - Mammalian Vav oncoprotein, a guanine-nucleotide exchange factor of the CDC24 family.
- Miscellanous proteins interacting with vertebrate receptor protein tyrosine kinases: oncoprotein Crk, mammalian cytoplasmic proteins Nck, Shc.
 - STAT proteins (signal transducers and activators of transcription).
 - Chicken tensin.
 - Yeast transcriptional control protein SPT6.

The profile developed to detect SH2 domains is based on a structural alignment consisting of 8 gap-free blocks and 7 linker regions totaling 92 match positions.

5

[1]

Sadowski I., Stone J.C., Pawson T.

Mol. Cell. Biol. 6:4396-4408(1986).

[2]

10 Russel R.B., Breed J., Barton G.J.

FEBS Lett. 304:15-20(1992).

[3]

Marangere L.E.M., Pawson T.

J. Cell Sci. Suppl. 18:97-104(1994).

15 [4]

Pawson T., Schlessinger J.

Curr. Biol. 3:434-442(1993).

[5]

Mayer B.J., Baltimore D.

20 Trends Cell. Biol. 3:8-13(1993).

[6]

Pawson T.

Nature 373:573-580(1995).

[7]

25 Kuriyan J., Cowburn D.

Curr. Opin. Struct. Biol. 3:828-837(1993).

824. Sulfate transporters signature

PROSITE cross-reference(s): PS01130; SULFATE_TRANSP

30

A number of proteins involved in the transport of sulfate across a membrane as well as some yet uncharacterized proteins have been shown [1,2] to be evolutionary related. These proteins are:

- Neurospora crassa sulfate permease II (gene cys-14).
- Yeast sulfate permeases (genes SUL1 and SUL2).
- Rat sulfate anion transporter 1 (SAT-1).
- Mammalian DTDST, a probable sulfate transporter which, in Human, is
- 5 involved in the genetic disease, diastrophic dysplasia (DTD).
 - Sulfate transporters 1, 2 and 3 from the legume Stylosanthes hamata.
 - Human pendrin (gene PDS), which is involved in a number of hearing loss genetic diseases.
- Human protein DRA (Down-Regulated in Adenoma).
 - Soybean early nodulin 70.
 - Escherichia coli hypothetical protein ychM.
 - Caenorhabditis elegans hypothetical protein F41D9.5.
- As expected by their transport function, these proteins are highly hydrophobic and seem to contain about 12 transmembrane domains. The best conserved region seems to be located in the second transmembrane region and is used as a signature pattern.
- Consensus pattern[PAV]-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVFYA]-[LIVST]-[YI]-x(3)-[GA]-[GST]-S-[KR]

[1]

Sandal N.N., Marcker K.A.

25 Trends Biochem. Sci. 19:19-19(1994).

[2]

Smith F.W., Hawkesford M.J., Prosser I.M., Clarkson D.T.

Mol. Gen. Genet. 247:709-715(1995).

30 825. TYA: TYA transposon protein

Ty are yeast transposons. A 5.7kb transcript codes for p3 a fusion protein of TYA and TYB. The TYA protein is analogous to the gag protein of retroviruses. TYA a is cleaved to form 46kd protein which can form mature virion like particles [1]. Number of members: 59

20

25

30

[1] Medline: 97404699. Cryo-electron microscopy structure of yeast Ty retrotransposon virus-like particles. Palmer KJ, Tichelaar W, Myers N, Burns NR, Butcher SJ, Kingsman AJ, Fuller SD, Saibil HR; J Virol 1997;71:6863-6868.

5 826. Aldolase II

Class II Aldolase and Adducin N-terminal domain.

-!- This family includes class II aldolases and adducins which have not been ascribed any enzymatic function. Number of members: 37

10 References:

- [1] Medline: 93294819. The spatial structure of the class II L-fuculose-1-phosphate aldolase from Escherichia coli. Dreyer MK, Schulz GE; J Mol Biol 1993;231:549-553.
- [2] Medline: 96256522. Catalytic mechanism of the metal-dependent fuculose aldolase from Escherichia coli as derived from the structure. Dreyer MK, Schulz GE; J Mol Biol 1996;259:458-466.

827. CBD 2

- -!- Two tryptophan residues are involved in cellulose binding.
- -!- Cellulose binding domain found in bacteria. Number of members: 51

References:

[1] Medline: 95284032. Solution structure of a cellulose-binding domain from Cellulomonas fimi by nuclear magnetic resonance spectroscopy. Xu GY, Ong E, Gilkes NR, Kilburn DG, Muhandiram DR, Harris-Brandts M, Carver JP, Kay LE, Harvey TS; Biochemistry 1995;34:6993-7009.

828. P

A unique feature of the eukaryotic subtilisin-like proprotein convertases is the presence of an additional highly conserved sequence of approximately 150 residues (P domain) located immediately downstream of the catalytic domain.

Number of members: 91

References:

30

[1] Medline: 94252314. A C-terminal domain conserved in precursor processing proteases is required for intramolecular N-terminal maturation of pro-Kex2 protease. Gluschankof P, Fuller RS; EMBO J 1994;13:2280-2288.

[2] Medline: 98225190. Regulatory roles of the P domain of the subtilisin-like prohormone convertases. Zhou A, Martin S, Lipkind G, LaMendola J, Steiner DF; J Biol Chem 1998;273:11107-11114.

829. Uncharacterized protein family UPF0020 signature PROSITE cross-reference(s): PS01261; UPF0020

The following uncharacterized proteins have been shown [1] to share regions of similarities:

- Escherichia coli hypothetical protein ycbY and HI0116/15, the corresponding Haemophilus influenzae protein.
- Bacillus subtilis hypothetical protein ypsC.
 - Synechocystis strain PCC 6803 hypothetical protein slr0064.
 - Methanococcus jannaschii hypothetical proteins MJ0438 and MJ0710.

These are hydrophilic proteins of from 40 Kd to about 80 Kd. They can be picked up in the database by the following pattern.

Consensus patternD-P-[LIVMF]-C-G-[ST]-G-x(3)-[LI]-E

References:

25 [1] Bairoch A. Unpublished observations (1997).

830. Uncharacterized protein family UPF0031 signatures
PROSITE cross-reference(s): PS01049; UPF0031_1; PS01050; UPF0031_2
The following uncharacterized proteins have been shown [1] to share regions of similarities:

- Yeast chromosome XI hypothetical protein YKL151c.
- Caenorhabditis elegans hypothetical protein R107.2.
- Escherichia coli hypothetical protein yjeF.

30

5

10

- Bacillus subtilis hypothetical protein yxkO.
- Helicobacter pylori hypothetical protein HP1363.
- Mycobacterium tuberculosis hypothetical protein MtCY77.05c.
- Mycobacterium leprae hypothetical protein B229 C2 201.
- Synechocystis strain PCC 6803 hypothetical protein sll1433.
- Methanococcus jannaschii hypothetical protein MJ1586.

These are proteins of about 30 to 40 Kd whose central region is well conserved. They can be picked up in the database by the following patterns.

Consensus pattern[SAV]-[IVW]-[LVA]-[LIV]-G-[PNS]-G-L-[GP]-x-[DENQT]
Consensus pattern[GA]-G-x-G-D-[TV]-[LT]-[STA]-G-x-[LIVM]

831. (ACOX)

15 Acyl-CoA oxidase

This is a family of Acyl-CoA oxidases EC:1.3.3.6. Acyl-coA oxidase converts acyl-CoA into trans-2-enoyl-CoA [1].

20 Number of members: 39

[1] Hayashi H, De Bellis L, Yamaguchi K, Kato A, Hayashi M, Nishimura M; Medline: 98192624. Molecular characterization of a glyoxysomal long chain acyl-CoA oxidase that is synthesized as a precursor of higher molecular mass in pumpkin." J Biol Chem 1998;273:8301-8307.

832. (AICARFT IMPCHas)

AICARFT/IMPCHase bienzyme

This is a family of bifunctional enzymes catalysing the last steps in de novo purine biosynthesis. The bifunctional enzyme is found in both prokaryotes and eukaryotes. The second last step is catalysed by 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase EC:2.1.2.3 (AICARFT), this enzyme catalyses the formylation of AICAR

15

20

25

30



with 10-formyl-tetrahydrofolate to yield FAICAR and tetrahydrofolate [1]. The last step is catalysed by IMP (Inosine monophosphate) cyclohydrolase EC:3.5.4.10 (IMPCHase), cyclizing FAICAR (5-formylaminoimidazole-4-carboxamide ribonucleotide) to IMP [1].

5 Number of members: 22

[1] Akira T, Komatsu M, Nango R, Tomooka A, Konaka K, Yamauchi M, Kitamura Y, Nomura S, Tsukamoto I; Medline: 97473523 Molecular cloning and expression of a rat cDNA encoding 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase" [published erratum appears in Gene 1998 Feb 27;208(2):337] Gene 1997;197:289-293.

[2] Rayl EA, Moroson BA, Beardsley GP; Medline: 96147205 The human purH gene product, 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase. Cloning, sequencing, expression, purification, kinetic analysis, and domain mapping." J Biol Chem 1996;271:2225-2233.

833. (AOX)

Alternative oxidase

The alternative oxi

The alternative oxidase is used as a second terminal oxidase in the mitochondria, electrons are transferred directly from reduced ubiquinol to oxygen forming water [2]. This is not coupled to ATP synthesis and is not inhibited by cyanide, this pathway is a single step process [1]. In rice the transcript levels of the alternative oxidase are increased by low temperature [1].

Number of members: 27

[1] Ito Y, Saisho D, Nakazono M, Tsutsumi N, Hirai A; Medline: 98086211 Transcript levels of tandem-arranged alternative oxidase genes in rice are increased by low temperature." Gene 1997;203:121-129.



[2] Li Q, Ritzel RG, McLean LL, McIntosh L, Ko T, Bertrand H, Nargang FE; Medline: 96366413 Cloning and analysis of the alternative oxidase gene of Neurospora crassa." Genetics 1996;142:129-140.

5

10

15

20

30

834. (APH)

Protein kinases signatures and profile

Cross-reference(s): PS00107; PROTEIN_KINASE_ATP, PS00108; PROTEIN_KINASE_ST, PS00109; PROTEIN_KINASE_TYR, PS50011; PROTEIN KINASE DOM

Eukaryotic protein kinases [1 to 5] are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. Two of these regions have been selected to build signature patterns. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme [6]; two signature patterns were derived for that region: one specific for serine/ threonine kinases and the other for tyrosine kinases. A profile was developed which is based on the alignment in [1] and covers the entire catalytic domain.

Consensus pattern: [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K [K binds ATP]

Sequences known to belong to this class detected by the pattern the majority of known protein kinases but it fails to find a number of them, especially viral kinases which are quite divergent in this region and are completely missed by this pattern.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3) [D is an active site residue]

Sequences known to belong to this class detected by the pattern. Most serine/ threonine specific protein kinases with 10 exceptions (half of them viral kinases) and also Epstein-Barr virus BGLF4 and Drosophila ninaC which have respectively Ser and Arg instead of the conserved Lys and which are therefore detected by the tyrosine kinase specific pattern described below.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3) [D is an active site residue] tyrosine specific protein kinases with the exception of human ERBB3 and mouse blk. This pattern will also detect most bacterial aminoglycoside phosphotransferases [8,9] and herpesviruses ganciclovir kinases [10]; which are proteins structurally and evolutionary related to protein kinases. Sequences known to belong to this class detected by the profile ALL, except for three viral kinases. This profile also detects receptor guanylate cyclases (see <PDOC00430>) and 2-5A-dependent ribonucleases.

Sequence similarities between these two families and the eukaryotic protein kinase family have been noticed before. It also detects Arabidopsis thaliana kinase-like protein TMKL1 which seems to have lost its catalytic activity.

Note if a protein analyzed includes the two protein kinase signatures, the probability of it being a protein kinase is close to 100%. Note eukaryotic-type protein kinases have also been found in prokaryotes such as Myxococcus xanthus [11] and Yersinia pseudotuberculosis. Note the patterns shown above has been updated since their publication in [7]. Note this documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so.

References

5

- [1] Hanks S.K., Hunter T., FASEB J. 9:576-596(1995).
- [2] Hunter T., Meth. Enzymol. 200:3-37(1991).
- 30 [3] Hanks S.K., Quinn A.M., Meth. Enzymol. 200:38-62(1991).
 - [4] Hanks S.K., Curr. Opin. Struct. Biol. 1:369-383(1991).
 - [5] Hanks S.K., Quinn A.M., Hunter T., Science 241:42-52(1988).
 - [6] Knighton D.R., Zheng J., Ten Eyck L.F., Ashford V.A., Xuong N.-H., Taylor, S.S., Sowadski J.M., Science 253:407-414(1991).

15

20

- [7] Bairoch A., Claverie J.-M., Nature 331:22(1988).
- [8] Benner S., Nature 329:21-21(1987).
- [9] Kirby R., J. Mol. Evol. 30:489-492(1992).
- [10] Littler E., Stuart A.D., Chee M.S., Nature 358:160-162(1992).
- 5 [11] Munoz-Dorado J., Inouye S., Inouye M., Cell 67:995-1006(1991).

835. (Asp_Glu_race)

Aspartate and glutamate racemases signatures

Cross-reference(s) PS00923; ASP_GLU_RACEMASE_1 PS00924; ASP_GLU_RACEMASE_2

Aspartate racemase (EC 5.1.1.13) and glutamate racemase (EC 5.1.1.3) are two evolutionary related bacterial enzymes that do not seem to require a cofactor for their activity [1]. Glutamate racemase, which interconverts L-glutamate into D-glutamate, is required for the biosynthesis of peptidoglycan and some peptide-based antibiotics such as gramicidin S. In addition to characterized aspartate and glutamate racemases, this family also includes a hypothetical protein from Erwinia carotovora and one from Escherichia coli (ygeA). Two conserved cysteines are present in the sequence of these enzymes. They are expected to play a role in catalytic activity by acting as bases in proton abstraction from the substrate. Signature patterns were developed for both cysteines.

Consensus pattern: [IVA]-[LIVM]-x-C-x(0,1)-N-[ST]-[MSA]-[STH]-[LIVFYSTANK]

Consensus pattern: [LIVM](2)-x-[AG]-C-T-[DEH]-[LIVMFY]-[PNGRS]-x-[LIVM]

[1] Gallo K.A., Knowles J.R., Biochemistry 32:3981-3990(1993).

836. (ATP-sulfurylase) ATP-sulfurylase

30

25

This family consists of ATP-sulfurylase or sulfate adenylyltransferase EC:2.7.7.4 some of which are part of a bifunctional polypeptide chain associated with adenosyl phosphosulphate (APS) kinase APS_kinase. Both enzymes are required for PAPS (phosphoadenosine-phosphosulfate) synthesis from inorganic sulphate [2]. ATP sulfurylase catalyses the synthesis of adenosine-phosphosulfate APS from ATP and inorganic sulphate [1].

Number of members: 37

- [1] Kurima K, Warman ML, Krishnan S, Domowicz M, Krueger RC Jr, Deyrup A, Schwartz
 NB; Medline: 98337975 A member of a family of sulfate-activating enzymes causes murine brachymorphism" [published erratum appears in Proc Natl Acad Sci U S A 1998 Sep 29;95(20):12071] Proc Natl Acad Sci U S A 1998;95:8681-8685.
- [2] Rosenthal E, Leustek T; Medline: 96096529 A multifunctional Urechis caupo protein,
 PAPS synthetase, has both ATP sulfurylase and APS kinase activities." Gene 1995;165:243-248.

837. (ATP-synt_F)

20 ATP synthase (F/14-kDa) subunit

This family includes 14-kDa subunit from vATPases [1], which is in the peripheral catalytic part of the complex [2]. The family also includes archaebacterial ATP synthase subunit F [3].

- Number of members: 23
 - [1] Guo Y, Kaiser K, Wieczorek H, Dow JA; Medline: 96269411 The Drosophila melanogaster gene vha14 encoding a 14-kDa F-subunit of the vacuolar ATPase." Gene 1996;172:239-243.
- [2] Peng SB, Crider BP, Tsai SJ, Xie XS, Stone DK; Medline: 96216416 Identification of a 14-kDa subunit associated with the catalytic sector of clathrin-coated vesicle H+-ATPase." J Biol Chem 1996;271:3324-3327.

[3] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968 Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon Methanosarcina mazei Go1." J Biol Chem 1996;271:18843-18852.

5

838. (CBD_4)

Starch binding domain

Number of members:

48

10

839. (CbiX)

15

The function of CbiX is uncertain, however it is found in cobalamin biosynthesis operons and so may have a related function. Some CbiX proteins contain a striking histidine-rich region at their C-terminus, which suggests that it might be involved in metal chelation [1].

Number of members: 6

20

[1] Raux E, Lanois A, Warren MJ, Rambach A, Thermes C; Medline: 98416126 Cobalamin (vitamin B12) biosynthesis: identification and characterization of a Bacillus megaterium cobI operon." Biochem J 1998;335:159-166.

25

840. (Complex1_51K)

Respiratory-chain NADH dehydrogenase 51 Kd subunit signatures Cross-reference(s) PS00644; COMPLEX1_51K_1 PS00645; COMPLEX1_51K_2

30

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 51 Kd (in mammals),

20

25

which is the second largest subunit of complex I and is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind to NAD, FMN, and a 2Fe-2S cluster.

The 51 Kd subunit is highly similar to [3,4]:

- 5 - Subunit alpha of Alcaligenes eutrophus NAD-reducing hydrogenase (gene hoxF) which also binds to NAD, FMN, and a 2Fe-2S cluster.
 - Subunit NQO1 of Paracoccus denitrificans NADH-ubiquinone oxidoreductase.
 - Subunit F of Escherichia coli NADH-ubiquinone oxidoreductase (gene nuoF).
- 10 The 51 Kd subunit and the bacterial hydrogenase alpha subunit contains three regions of sequence similarities. The first one most probably corresponds to the NAD-binding site, the second to the FMN-binding site, and the third one, which contains three cysteines, to the ironsulfur binding region. Signature patterns have been developed for the FMN-binding and for the 2Fe-2S binding regions.

Consensus pattern: G-[AM]-G-[AR]-Y-[LIVM]-C-G-[DE](2)-[STA](2)-[LIM](2)-[EN]-S Consensus pattern: E-S-C-G-x-C-x-P-C-R-x-G [The three C's are putative 2Fe-2S ligands]

- [1] Ragan C.I., Curr. Top. Bioenerg. 15:1-36(1987).
- [2] Weiss H., Friedrich T., Hofhaus G., Preis D., Eur. J. Biochem. 197:563-576(1991).
 - [3] Fearnley I.M., Walker J.E. Biochim. Biophys. Acta 1140:105-134(1992).
 - [4] Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H., J. Mol. Biol. 233:109-122(1993).

841. (DAP epimerase)

Diaminopimelate epimerase signature

Cross-reference(s) PS01326; DAP EPIMERASE

Diaminopimelate epimerase (EC 5.1.1.7) catalyzes the isomeriazation of L,L- to D,L-meso-30 diaminopimelate in the biosynthetic pathway leading from aspartate to lysine. This enzyme is a protein of about 30 Kd. Two conserved cysteines seem [1] to function as the acid and base in the catalytic mechanism. As a signature pattern, the region surrounding the first of these two active site cysteines were selected.

20

5

Consensus pattern: N-x-D-G-S-x(4)-C-G-N-[GA]-x-R [C is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for an Anabaena dapF which has a Ser instead of the active site Cys.

[1] Cirilli M., Zheng R., Scapin G., Blanchard J.S., Biochemistry 37:16452-16458(1998).

842. (DNA gyraseB C)

10 DNA topoisomerase II signature

Cross-reference(s) PS00177; TOPOISOMERASE II

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaebacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaebacteria the enzyme, known as DNA gyrase, consists of two subunits (genes gyrA and gyrB [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes parC and parE). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

<-------About-1400-residues----->

| The second of the second of

ASF

15

20

30

5

As a signature pattern for this family of proteins, a region that contains a highly conserved pentapeptide was selected. The pattern is located in gyrB, in parE, and in protein 39 of phage T4 topoisomerase.

Consensus pattern: [LIVMA]-x-E-G-[DN]-S-A-x-[STAG]

- 10 [1] Sternglanz R., Curr. Opin. Cell Biol. 1:533-535(1990).
 - [2] Bjornsti M.-A., Curr. Opin. Struct. Biol. 1:99-103(1991).
 - [3] Sharma A., Mondragon A., Curr. Opin. Struct. Biol. 5:39-47(1995).
 - [4] Roca J., Trends Biochem. Sci. 20:156-160(1995).

843. (DUF16)

Protein of unknown function

The function of this protein is unknown. It appears to only occur in Mycoplasma pneumoniae.

Number of members: 26

[1] Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R; Medline: 97105885 Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae." Nucleic Acids Res 1996;24:4420-4449.

844. (DUF21)

Domain of unknown function

This transmembrane region has no known function. Many of the sequences in this family are annotated as hemolysins, however this is due to a similarity to Swiss:Q54318 that does not

10

15

20

25

30

847. (FF)

FF domain

685

contain this domain. This domain is found in the N-terminus of the proteins adjacent to two intracellular CBS domains CBS.

42

845. (DUF56)

Number of members:

Integral membrane protein

The members of this family are putative integral membrane proteins. The function of the family is unknown, however the family includes Sec59 from yeast. Sec59 is a dolichol kinase EC:2.7.1.108, but it is not clear if the enzymatic activity resides in this region or its N

terminal region.

Number of members: 13

846. (DUF94)

Domain of unknown function

Number of members: 9

The function of this domain is unknown. It is found in both eukaryotes and archaebacteria. The alignment contains a completely conserved aspartate residue that may be functionally important. The eukaryotic domains contains three conserved cysteines and a histidine that

might be metal binding, however these are absent in the archaebacterial proteins.

20

30

686

This domain may be involved in protein-protein interaction [1].

Number of members: 42

5 [1] Bedford MT, Leder P; Medline: 99322199 The FF domain: a novel motif that often accompanies WW domains." Trends Biochem Sci 1999;24:264-265.

848. (FLO_LFY)

10 Floricaula / Leafy protein

This family consists of various plant development proteins which are homologues of floricaula (FLO) and Leafy (LFY) proteins which are floral meristem identity proteins. Mutations in the sequences of these proteins affect flower and leaf development.

Number of members: 16

[1] Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, Ellis N; Medline: 97411151 UNIFOLIATA regulates leaf and flower morphogenesis in pea." Curr Biol 1997;7:581-587.

[2] Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM; Medline: 92274452 LEAFY controls floral meristem identity in Arabidopsis." Cell 1992;69:843-859.

25 **849.** (G-patch)

G-patch domain

This domain is found in a number of RNA binding proteins, and is also found in proteins that contain RNA binding domains. This suggests that this domain may have an RNA binding function. This domain has seven highly conserved glycines.

Number of members: 47

[1] Aravind L, Koonin EV; Medline: 10470032 G-patch: a new conserved domain in

eukaryotic RNA-processing proteins and type D retroviral polyproteins." Trends Biochem Sci 1999;24:342-344.

5

10

15

20

30

850. (Gram-ve_porins)

General diffusion Gram-negative porins signature

Cross-reference(s) PS00576; GRAM_NEG_PORIN

The outer membrane of Gram-negative bacteria acts as a molecular filter for hydrophilic compounds. Proteins, known as porins [1], are responsible for the 'molecular sieve' properties of the outer membrane. Porins form large water- filled channels which allows the diffusion of hydrophilic molecules into the periplasmic space. Some porins form general diffusion channels that allows any solutes up to a certain size (that size is known as the exclusion limit) to cross the membrane, while other porins are specific for a solute and contain a binding site for that solute inside the pores (these are known as selective porins). As porins are the major outer membrane proteins, they also serve as receptor sites for the binding of phages and bacteriocins. General diffusion porins generally assemble as trimer in the membrane and the transmembrane core of these proteins is composed exclusively of beta strands [2]. It has been shown [3] that a number of general porins are evolutionary related, these porins are:

- Enterobacteria phoE.
- Enterobacteria ompC.
- Enterobacteria ompF.
- Enterobacteria nmpC.
- Bacteriophage PA-2 LC.
 - Neisseria PI.A.
 - Neisseria PI.B.

As a signature pattern a conserved region was selected, located in the C-terminal part of these proteins, which spans two putative transmembrane beta strands.

Consensus pattern: [LIVMFY]-x(2)-G-x(2)-Y-x-F-x-K-x(2)-[SN]-[STAV]-[LIVMFYW]- V

[1] Benz R., Bauer K., Eur. J. Biochem. 176:1-19(1988).

15

20

[2] Jap B.K., Walian P.J., Q. Rev. Biophys. 23:367-403(1990).

[3] Jeanteur D., Lakey J.H., Pattus F., Mol. Microbiol. 5:2153-2164(1991).

5 **851. (HlyD)**

HlyD family secretion proteins signature

Cross-reference(s) PS00543; HLYD_FAMILY

Gram-negative bacteria produce a number of proteins which are secreted into the growth medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, require the help of two or more proteins for their secretion across the cell envelope. Amongst which a protein belonging to the ABC transporters family (see the relevant entry <PDOC00185>) and a protein belonging to a family which is currently composed [1 to 5] of the following members:

Gene Species Protein which is exported

hlyD Escherichia coli Hemolysin

appD A.pleuropneumoniae Hemolysin

lcnD Lactococcus lactis Lactococcin A

lktD A.actinomycetemcomitans Leukotoxin

Pasteurella haemolytica

rtxD A.pleuropneumoniae Toxin-III

cvaD Bordetella pertussis Calmodulin-sensitive adenylate cyclase-

hemolysin (cyclolysin)

25 cvaA Escherichia coli Colicin V

prtE Erwinia chrysanthemi Extracellular proteases B and C

aprE Pseudomonas aeruginosa Alkaline protease

emrA Escherichia coli Drugs and toxins

yicR Escherichia coli Unknewn

These proteins are evolutionary related and consist of from 390 to 480 amino acid residues. They seem to be anchored in the inner membrane by a N-terminal transmembrane region. Their exact role in the secretion process is not yet known. The C-terminal section of these proteins is the best conserved region; a signature pattern from that region was derived.



Consensus pattern: [LIVM]-x(2)-G-[LM]-x(3)-[STGAV]-x-[LIVMT]-x-[LIVMT]-[GE]-x-[KR]-x-[LIVMFYW](2)-x-[LIVMFYW](3)

Sequences known to belong to this class detected by the pattern ALL, except for emrA and yjcR.

5

References:

- [1] Gilson L., Mahanty H.K., Kolter R., EMBO J. 9:3875-3884(1990).
- [2] Letoffe S., Delepelaire P., Wandersman C., EMBO J. 9:1375-1382(1990).
- [3] Stoddard G.W., Petzel J.P., van Belkum M.J., Kok J., McKay L.L., Appl. Environ.
- 10 Microbiol. 58:1952-1961(1992).
 - [4] Duong F., Lazdunski A., Cami B., Murgier M., Gene 121:47-54(1992).
 - [5] Lewis K., Trends Biochem. Sci. 19:119-123(1994).

15 852. (IBR)

In Between Ring fingers

The IBR (In Between Ring fingers) domain is found to occur between pairs of ring fingers (zf-C3HC4). The function of this domain is unknown. This domain has also been called the C6HC domain and DRIL (for double RING finger linked) domain [2].

Number of members: 25

- [1] Morett E, Bork P; Medline: 10366851 A novel transactivation domain in parkin."Trends Biochem Sci 1999;24:229-231.
- [2] van der Reijden BA, Erpelinck-Verschueren CA, Lowenberg B, Jansen JH; Medline: 99349709 TRIADs: a new class of proteins with a novel cysteine-rich signature." Protein Sci 1999;8:1557-1561.
- 30 **853. (IPPT)**

IPP transferase

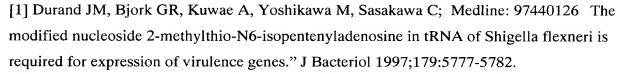
10

15

20

25

30



- [2] Boguta M, Hunter LA, Shen WC, Gillman EC, Martin NC, Hopper AK; Medline:
- 94187700 Subcellular locations of MOD5 proteins: mapping of sequences sufficient for targeting to mitochondria and demonstration that mitochondrial and nuclear isoforms commingle in the cytosol." Mol Cell Biol 1994;14:2298-2306.
 - [3] Gillman EC, Slusher LB, Martin NC, Hopper AK; Medline: 91203856 MOD5 translation initiation sites determine N6-isopentenyladenosine modification of mitochondrial and cytoplasmic tRNA." Mol Cell Biol 1991;11:2382-2390.

854. (KE2)

KE2 family protein

The function of members of this family is unknown, although they have been suggested to contain a DNA binding leucine zipper motif [2].

Number of members: 9

- [1] Ha H, Abe K, Artzt K; Medline: 92084131 Primary structure of the embryo-expressed gene KE2 from the mouse H-2K region." Gene 1991;107:345-346.
- [2] Shang HS, Wong SM, Tan HM, Wu M; Medline: 95129859 YKE2, a yeast nuclear gene encoding a protein showing homology to mouse KE2 and containing a putative leucine-zipper motif." Gene 1994;151:197-201.
- ,

855. (Lipoprotein 6)

Prokaryotic membrane lipoprotein lipid attachment site

Cross-reference(s) PS00013; PROKAR LIPOPROTEIN

In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which

- a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]):
- Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp).
- Escherichia coli lipoprotein-28 (gene nlpA).
- 5 Escherichia coli lipoprotein-34 (gene nlpB).
 - Escherichia coli lipoprotein nlpC.
 - Escherichia coli lipoprotein nlpD.
 - Escherichia coli osmotically inducible lipoprotein B (gene osmB).
 - Escherichia coli osmotically inducible lipoprotein E (gene osmE).
- Escherichia coli peptidoglycan-associated lipoprotein (gene pal).
 - Escherichia coli rare lipoproteins A and B (genes rplA and rplB).
 - Escherichia coli copper homeostasis protein cutF (or nlpE).
 - Escherichia coli plasmids traT proteins.
 - Escherichia coli Col plasmids lysis proteins.
- A number of Bacillus beta-lactamases.
 - Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA).
 - Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB).
 - Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7).
 - Chlamydia trachomatis outer membrane protein 3 (gene omp3).
- Fibrobacter succinogenes endoglucanase cel-3.
 - Haemophilus influenzae proteins Pal and Pcp.
 - Klebsiella pullulunase (gene pulA).
 - Klebsiella pullulunase secretion protein pulS.
 - Mycoplasma hyorhinis protein p37.
- Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlpABC).
 - Neisseria outer membrane protein H.8.
 - Pseudomonas aeruginosa lipopeptide (gene lppL).
 - Pseudomonas solanacearum endoglucanase egl.
 - Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC).
- Rickettsia 17 Kd antigen.
 - Shigella flexneri invasion plasmid proteins mxiJ and mxiM.
 - Streptococcus pneumoniae oligopeptide transport protein A (gene amiA).
 - Treponema pallidium 34 Kd antigen.
 - Treponema pallidium membrane protein A (gene tmpA).

20

25

30

5



- Yersinia virulence plasmid protein yscJ.
- Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper-binding protein. This is the first archaebacterial protein known to be modified in such a fashion).

From the precursor sequences of all these proteins, a consensus pattern and a set of rules to identify this type of post-translational modification were derived.

Consensus pattern: {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1)

The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be.

References

- [1] Hayashi S., Wu H.C., J. Bioenerg. Biomembr. 22:451-471(1990).
- [2] Klein P., Somorjai R.L., Lau P.C.K., Protein Eng. 2:15-20(1988).
- [3] von Heijne G., Protein Eng. 2:531-534(1989).
- [4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

856. (Lipoprotein_7)

Adhesin lipoprotein

This family consists of the p50 and variable adherence-associated antigen (Vaa) adhesins from Mycoplasma hominis. M. hominis is a mycoplasma associated with human urogenital diseases, pneumonia, and septic arthritis [1]. An adhesin is a cell surface molecule that mediates adhesion to other cells or to the surrounding surface or substrate. The Vaa antigen is a 50-kDa surface lipoprotein that has four tandem repetitive DNA sequences encoding a periodic peptide structure, and is highly immunogenic in the human host [1]. p50 is also a 50-

kDa lipoprotein, having three repeats A,B and C, that may be a tetramer of 191-kDa in its native environment [2].

Number of members: 18

5

- [1] Zhang Q, Wise KS; Medline: 96294788 Molecular basis of size and antigenic variation of a Mycoplasma hominis adhesin encoded by divergent vaa genes. Infect Immun 1996;64:2737-2744.
- [2] Henrich B, Kitzerow A, Feldmann RC, Schaal H, Hadding U; Medline: 97047675

 Repetitive elements of the Mycoplasma hominis adhesin p50 can be differentiated by monoclonal antibodies." Infect Immun 1996;64:4027-4034.

857. (MaoC like)

15 MaoC like domain

The MaoC protein is found to share similarity with a wide variety of enzymes; estradiol 17 beta-dehydrogenase 4, peroxisomal hydratase-dehydrogenase-epimerase, fatty acid synthase beta subunit. All these enzymes contain other domains. This domain is also present in the NodN nodulation protein N. No specific function has been assigned to this region of any of these proteins. The maoC gene is part of a operon with maoA which is involved in the synthesis of monoamine oxidase [1].

Number of members: 46

25

20

[1] Sugino H, Sasaki M, Azakami H, Yamashita M, Murooka Y Medline: 96235221 A monoamine-regulated Klebsiella aerogenes operon containing the monoamine oxidase structural gene (maoA) and the maoC gene." J Bacteriol 1992;174:2485-2492.

30

858. (MSP)

Manganese-stabilizing protein / photosystem II polypeptide

This family consists of the 33 KDa photosystem II polypeptide from the oxygen evolving complex (OEC) of plants and cyanobacteria. The protein is also known as the manganesestabilizing protein as it is associated with the manganese complex of the OEC and may provide the ligands for the complex [1].

5

Number of members: 17

[1] Philbrick JB, Zilinskas BA; Medline: 88334494 "Cloning, nucleotide sequence and mutational analysis of the gene encoding the Photosystem II manganese-stabilizing polypeptide of Synechocystis 6803." Mol Gen Genet 1988;212:418-425.

10

859. (NAC)

15

[1] Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV; Medline: 99342100 Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell." Genome Res 1999;9:608-628.

20

Number of members:

27

860. (Nop)

Putative snoRNA binding domain

25

30

This family consists of various Pre RNA processing ribonucleoproteins. The function of the aligned region is unknown however it may be a common RNA or snoRNA or Nop1p binding domain. Nop5p (Nop58p) Swiss:Q12499 from yeast is the protein component of a ribonucleoprotein protein required for pre-18s rRNA processing and is suggested to function with Nop1p in a snoRNA complex [1]. Nop56p Swiss:O00567 and Nop5p interact with Nop1p and are required for ribosome biogenesis [2]. Prp31p Swiss:p49704 is required for pre-mRNA splicing in S. cerevisiae [3].

Number of members:

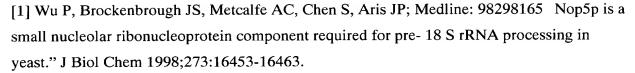
23

15

20

25

30



- [2] Gautier T, Berges T, Tollervey D, Hurt E; Medline: 8038777 Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis." Mol Cell Biol 1997;17:7088-7098.
 - [3] Weidenhammer EM, Singh M, Ruiz-Noriega M, Woolford JL Jr; Medline: 96184869

 The PRP31 gene encodes a novel protein required for pre-mRNA splicing in Saccharomyces cerevisiae." Nucleic Acids Res 1996;24:1164-1170.

861. (Nramp)

Natural resistance-associated macrophage protein

The natural resistance-associated macrophage protein (NRAMP) family consists of Nramp1, Nramp2, and yeast proteins Smf1 and Smf2. The NRAMP family is a novel family of functional related proteins defined by a conserved hydrophobic core of ten transmembrane domains [5]. This family of membrane proteins are divalent cation transporters. Nramp1 is an integral membrane protein expressed exclusively in cells of the immune system and is recruited to the membrane of a phagosome upon phagocytosis [1]. By controlling divalent cation concentrations Nramp1 may regulate the interphagosomal replication of bacteria [1]. Mutations in Nramp1 may genetically predispose an individual to susceptibility to diseases including leprosy and tuberculosis conversely this might however provide protection form rheumatoid arthritis [1]. Nramp2 is a multiple divalent cation transporter for Fe2+, Mn2+ and Zn2+ amongst others it is expressed at high levels in the intestine; and is major transferrinindependent iron uptake system in mammals [1]. The yeast proteins Smf1 and Smf2 may also transport divalent cations [3].

Number of members: 36

[1] Govoni G, Gros P; Medline: 98383996 Macrophage NRAMP1 and its role in resistance to microbial infections." Inflamm Res 1998;47:277-284.

20

30

5

- [2] Agranoff DD, Krishna S Medline: 98294035 Metal ion homeostasis and intracellular parasitism." Mol Microbiol 1998;28:403-412.
- [3] Pinner E, Gruenheid S, Raymond M, Gros P; Medline: 98030569 Functional complementation of the yeast divalent cation transporter family SMF by NRAMP2, a member of the mammalian natural resistance- associated macrophage protein family." J Biol Chem 1997;272:28933-28938.
- [4] Cellier M, Belouchi A, Gros P; Medline: 96402487 Resistance to intracellular infections: comparative genomic analysis of Nramp." Trends Genet 1996;12:201-204.
- [5] Cellier M, Prive G, Belouchi A, Kwan T, Rodrigues V, Chia W, Gros P; Medline:
 96036029 Nramp defines a family of membrane proteins." Proc Natl Acad Sci U S A
 1995;92:10089-10093.

862. (NTP transf 2)

Nucleotidyltransferase domain

Members of this family belong to a large family of nucleotidyltransferases [1].

Number of members: 83

- [1] Holm L, Sander C; Medline: 96005605 DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily." Trends Biochem Sci 1995;20:345-347.
- 25 863. (Paramyxo_P)
 Paramyxovirus P phosphoprotein

This family consists of paramyxovirus P phosphoprotein from sendai virus and human and bovine parainfluenza viruses. The P protein is an essential part of the viral RNA polymerase complex formed form the P and L proteins [1]. The exact role of the P protein in this complex in unknown but it is involved in multiple protein-protein interactions and binding the polymerase complex to the nucleocapsid or ribonucleoprotein template [1]. It also appears to be important for the proper folding of the L protein [1]. The paramyxoviruses have a negative sense ssRNA genome [1].

cysteine-rich V protein." J Virol 1991;65:3406-3410.

21

15

Number of members:

[1] Bowman MC, Smallwood S, Moyer SA; Medline: 99329169 Dissection of Individual Functions of the Sendai Virus Phosphoprotein in Transcription." J Virol 1999;73:6474-6483.
[2] Matsuoka Y, Curran J, Pelet T, Kolakofsky D, Ray R, Compans RW; Medline: 91237868 The P gene of human parainfluenza virus type 1 encodes P and C proteins but not a

10

5

864. (Patatin)

This family consists of various patatin glycoproteins from plants. The patatin protein accounts for up to 40% of the total soluble protein in potato tubers [2]. Patatin is a storage protein but it also has the enzymatic activity of lipid acyl hydrolase, catalysing the cleavage of fatty acids from membrane lipids [2].

Number of members:

[1] Banfalvi Z, Kostyal Z, Barta E; Medline: 95107249 Solanum brevidens possesses a non-sucrose-inducible patatin gene." Mol Gen Genet 1994;245:517-522.

[2] Mignery GA, Pikaard CS, Park WD; Medline: 88226014 Molecular characterization of the patatin multigene family of potato." Gene 1988;62:27-44.

25

865. (Pentapeptide 2)

Pentapeptide repeats (8 copies)

These repeats are found in many mycobacterial proteins. These repeats are most common in the PPE family of proteins, where they are found in the MPTR subfamily of PPE proteins. The function of these repeats is unknown. The repeat can be approximately described as XNXGX, where X can be any amino acid. These repeats are similar to Pentapeptide [1], however it is not clear if these two families are structurally related.

5

10

15

20

25

30

Number of members:

[1] Bateman A, Murzin A, Teichmann SA; Medline: 98318059 Structure and distribution of pentapeptide repeats in bacteria." Protein Sci 1998;7:1477-1480.

[2] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG; Medline: 98295987 Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence." Nature 1998;393:537-544.

866. (Peptidase_C13)

Peptidase C13 family

This family of peptidases is known as the hemoglobinase family because it contains a globin degrading enzyme from blood parasites Swiss:P42665. However relatives are found in plants and other organisms that have other functions. Members of this family are asparaginyl peptidases [1].

Number of members: 26

[1] Chen JM, Dando PM, Rawlings ND, Brown MA, Young NE, Stevens RA, Hewitt E, Watts C, Barrett AJ; Medline: 97218252 Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase." J Biol Chem 1997;272:8090-8098.

867. (Pro dh)

Proline dehydrogenase

Number of members:

25

[1] Ling M, Allen SW, Wood JM; Medline: 95055736 Sequence analysis identifies the proline dehydrogenase and delta 1- pyrroline-5-carboxylate dehydrogenase domains of the multifunctional Escherichia coli PutA protein." J Mol Biol 1994;243:950-956.

5

10

15

20

30

868. (PsbP)

This family consists of the 23 kDa subunit of oxygen evolving system of photosystem II or PsbP from various plants (where it is encoded by the nuclear genome) and Cyanobacteria. The 23 KDa PsbP protein is required for PSII to be fully operational in vivo, it increases the affinity of the water oxidation site for Cl- and provides the conditions required for high affinity binding of Ca2+ [2].

Number of members:

25

[1] Rova EM, Mc Ewen B, Fredriksson PO, Styring S; Medline: 97067138 Photoactivation and photoinhibition are competing in a mutant of Chlamydomonas reinhardtii lacking the 23-kDa extrinsic subunit of photosystem II." J Biol Chem 1996;271:28918-28924.

[2] Kochhar A, Khurana JP, Tyagi AK; Medline: 97191538 Nucleotide sequence of the psbP gene encoding precursor of 23-kDa polypeptide of oxygen-evolving complex in Arabidopsis thaliana and its expression in the wild-type and a constitutively photomorphogenic mutant." DNA Res 1996;3:277-285.

25 **869. (PUA)**

The PUA domain named after PseudoUridine synthase and Archaeosine transglycosylase, was detected in archaeal and eukaryotic pseudouridine synthases, archaeal archaeosine synthases, a family of predicted ATPases that may be involved in RNA modification, a family of predicted archaeal and bacterial rRNA methylases. Additionally, the PUA domain was detected in a family of eukaryotic proteins that also contain a domain homologous to the translation initiation factor eIF1/SUI1; these proteins may comprise a novel type of translation factors. Unexpectedly, the PUA domain was detected also in bacterial and yeast glutamate kinases; this is compatible with the demonstrated role of these enzymes in the

regulation of the expression of other genes [1]. It is predicted that the PUA domain is an RNA binding domain.

Number of members: 48

5

- [1] Aravind L, Koonin EV; Medline: 99193178 Novel predicted RNA-binding domains associated with the translation machinery." J Mol Evol 1999;48:291-302.
- 10 870. (RF1) eRF1-like proteins

Members of this family are peptide chain release factors. The eukaryotic Release Factor 1 proteins (eRF1s) are involved in termination of translation. The eRF1 protein is functional for all stop codons and appears to abolish read-through of these codons. This family also includes other proteins for which the precise molecular function is unknown. Many of them are from Archaebacteria. These proteins may also be involved in translation termination but this awaits experimental verification. Number of members: 25

- [1] Frolova L, Le Goff X, Rasmussen HH, Cheperegin S, Drugeon G, Kress M, Arman I, Haenni AL, Celis JE, Philippe M, et al; Medline: 95082951 A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor" [see comments] Nature 1994;372:701-703.
- [2] Drugeon G, Jean-Jean O, Frolova L, Le Goff X, Philippe M, Kisselev L, Haenni AL;
 Medline: 97315314 Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all three termination codons in messenger RNA." Nucleic Acids Res 1997;25:2254-2258.
- 30 871. (Ribosomal_L14e)Ribosomal protein L14

 This family includes the eukaryotic ribosomal protein L14.

 Number of members: 15

10

15

20

872. (Ribosomal_S27)

Ribosomal protein S27a

This family of ribosomal proteins consists mainly of the 40S ribosomal protein S27a which is synthesized as a C-terminal extension of ubiquitin (CEP). The S27a domain compromises the C-terminal half of the protein. The synthesis of ribosomal proteins as extensions of ubiquitin promotes their incorporation into nascent ribosomes by a transient metabolic stabilization and is required for efficient ribosome biogenesis [3]. The ribosomal extension protein S27a contains a basic region that is proposed to form a zinc finger; its fusion gene is proposed as a mechanism to maintain a fixed ratio between ubiquitin necessary for degrading proteins and ribosomes a source of proteins [2].

Number of members: 36

873. (Spermine_synth)

Spermine/spermidine synthase

Spermine and spermidine are polyamines. This family includes spermidine synthase that catalyses the fifth (last) step in the biosynthesis of spermidine from arginine, and spermine synthase.

Number of members: 39

- [1] Mezquita J, Pau M, Mezquita C; Medline: 97449308 Characterization and expression of two chicken cDNAs encoding ubiquitin fused to ribosomal proteins of 52 and 80 amino acids." Gene 1997;195:313-319.
 - [2] Redman KL, Rechsteiner M; Medline: 89181932 Identification of the long ubiquitin extension as ribosomal protein S27a." Nature 1989;338:438-440.
- [3] Finley D, Bartel B, Varshavsky A; Medline: 89181925 The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis." Nature 1989;338:394-401.

874. (Surp)

Surp module

5

10

[1] Denhez F, Lafyatis R; Medline: 94266805 Conservation of regulated alternative splicing and identification of functional domains in vertebrate homologs to the Drosophila splicing regulator, suppressor-of-white-apricot." J Biol Chem 1994;269:16170-16179.

This domain is also known as the SWAP domain. SWAP stands for Suppressor-of-White-APricot. It has been suggested that these domains may be RNA binding [1].

Number of members:

32

875. (TFIIE)

15 TFIIE alpha subunit

The general transcription factor TFIIE has an essential role in eukaryotic transcription initiation together with RNA polymerase II and other general factors. Human TFIIE consists of two subunits TFIIE-alpha Swiss:P29083 and TFIIE-beta Swiss:P29084 and joins the preinitiation complex after RNA polymerase II and TFIIF [1]. This family consists of the conserved amino terminal region of eukaryotic TFIIE-alpha [2] and proteins from archaebacteria that are presumed to be TFIIE-alpha subunits also Swiss:O29501 [3].

Number of members: 12

25

30

20

- [1] Ohkuma Y, Sumimoto H, Hoffmann A, Shimasaki S, Horikoshi M, Roeder RG; Medline: 92065982 Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE." Nature 1991;354:398-401.
- [2] Ohkuma Y, Hashimoto S, Roeder RG, Horikoshi M; Medline: 93087200 Identification of two large subdomains in TFIIE-alpha on the basis of homology between Xenopus and human sequences. Nucleic Acids Res 1992;20:5838-5838.
- [3] Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA,

30

McKenney K, Adams MD, Loftus B, Venter JC, et al; Medline: 98049343 The complete genome sequence of the hyperthermophilic, sulphate- reducing archaeon Archaeoglobus fulgidus." Nature 1997;390:364-370.

5

876. (Transglut_core)

Cross-reference(s) PS00547; TRANSGLUTAMINASES

- 10 Transglutaminases (EC 2.3.2.13) (TGase) [1,2] are calcium-dependent enzymes that catalyze the cross-linking of proteins by promoting the formation of isopeptide bonds between the gamma-carboxyl group of a glutamine in one polypeptide chain and the epsilon-amino group of a lysine in a second polypeptide chain. TGases also catalyze the conjugation of polyamines to proteins. The best known transglutaminase is blood coagulation factor XIII, a plasma tetrameric protein composed of two catalytic A subunits and two non-catalytic B subunits. Factor XIII is responsible for cross-linking fibrin chains, thus stabilizing the fibrin clot. Other forms of transglutaminases are widely distributed in various organs, tissues and body fluids. Sequence data is available for the following forms of TGase:
 - Transglutaminase K (Tgase K), a membrane-bound enzyme found in mammalian epidermis and important for the formation of the cornified cell envelope (gene TGM1).
 - Tissue transglutaminase (TGase C), a monomeric ubiquitous enzyme located in the cytoplasm (gene TGM2).
 - Transglutaminase 3, responsible for the later stages of cell envelope formation in the epidermis and the hair follicle (gene TGM3).
- Transglutaminase 4 (gene TGM4).

A conserved cysteine is known to be involved in the catalytic mechanism of TGases. The erythrocyte membrane band 4.2 protein, which probably plays an important role in regulating the shape of erythrocytes and their mechanical properties, is evolutionary related to TGases. However the active site cysteine is substituted by an alanine and the 4.2 protein does not

show TGase activity.

20

25

Consensus pattern:[GT]-Q-[CA]-W-V-x-[SA]-[GA]-[IVT]-x(2)-T-x-[LMSC]-R-[CSA]-[LV]-G [The first C is the active site residue] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Ichinose A., Bottenus R.E., Davie E.W. J. Biol. Chem. 265:13411-13414(1990).
 [2] Greenberg C.S., Birckbichler P.J., Rice R.H. FASEB J. 5:3071-3077(1991).

877. (TruB_N)

10 TruB family pseudouridylate synthase (N terminal domain)

Members of this family are involved in modifying bases in RNA molecules. They carry out the conversion of uracil bases to pseudouridine. This family includes TruB, a pseudouridylate synthase that specifically converts uracil 55 to pseudouridine in most tRNAs. This family also includes Cbf5p that modifies rRNA [2].

Number of members: 33

[1] Nurse K, Wrzesinski J, Bakin A, Lane BG, Ofengand J; Medline: 96079944 Purification, cloning, and properties of the tRNA psi 55 synthase from Escherichia coli." RNA 1995;1:102-112.

[2] Lafontaine DLJ, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D; Medline: 98139521 The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase." Genes Dev 1998;12:527-537.

878. (UDPGP)

UTP--glucose-1-phosphate uridylyltransferase

This family consists of UTP--glucose-1-phosphate uridylyltransferases, EC:2.7.7.9. Also known as UDP-glucose pyrophosphorylase (UDPGP) and Glucose-1-phosphate uridylyltransferase. UTP--glucose-1-phosphate uridylyltransferase catalyses the interconversion of MgUTP + glucose-1-phosphate and UDP-glucose + MgPPi [1]. UDP-glucose is an important intermediate in mammalian carbohydrate interconversion involved in

various metabolic roles depending on tissue type [1]. In Dictyostelium (slime mold) mutants in this enzyme abort the development cycle [2]. Also within the family is UDP-N-acetylglucosamine Swiss:Q16222 or AGX1 [3] and two hypothetical proteins from Borrelia burgdorferi the lyme disease spirochaete Swiss:O51893 and Swiss:O51036.

Number of members:

5

10

s: 18

[1] Duggleby RG, Chao YC, Huang JG, Peng HL, Chang HY; Medline: 96202932 Sequence differences between human muscle and liver cDNAs for UDPglucose pyrophosphorylase and kinetic properties of the recombinant enzymes expressed in Escherichia coli." Eur J Biochem 1996;235:173-179.

- [2] Ragheb JA, Dottin RP; Medline: 87231075 Structure and sequence of a UDP glucose pyrophosphorylase gene of Dictyostelium discoideum." Nucleic Acids Res 1987;15:3891-3906.
- [3] Mio T, Yabe T, Arisawa M, Yamada-Okabe H; Medline: 98269105 The eukaryotic UDP-N-acetylglucosamine pyrophosphorylases. Gene cloning, protein expression, and catalytic mechanism. J Biol Chem 1998;273:14392-14397.

20 **879. (UPF004)**

Uncharacterized protein family UPF0044 signature Cross-reference(s) PS01301; UPF0044

The following uncharacterized proteins have been shown [1] to be highly similar:

- Bacillus subtilis hypothetical protein yqeI.
 - Escherichia coli hypothetical protein yhbY and HI1333, the corresponding Haemophilus influenzae protein.
 - Methanococcus jannaschii hypothetical protein MJ0652.

These are small proteins of 10 to 15 Kd. They can be picked up in the database by the following pattern. This pattern is located in the N-terminal part of these proteins.

25

30

706

Consensus pattern: L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)- [LIV]-[GA]-x(2)-G Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

880. (zf-A20)

A20-like zinc finger

A20- (an inhibitor of cell death)-like zinc fingers. The zinc finger mediates self-association in A20. These fingers also

mediate IL-1-induced NF-kappa B activation.

Number of members: 22

- [1] Heyninck K, Beyaert R; Medline: 99126071 The cytokine-inducible zinc finger protein A20 inhibits IL-1-induced NF- kappaB activation at the level of TRAF6. FEBS Lett 1999;442:147-150.
- [2] De Valck D, Heyninck K, Van Criekinge W, Contreras R, Beyaert R, Fiers W; Medline: 96390831 A20, an inhibitor of cell death, self-associates by its zinc finger domain." FEBS Lett 1996;384:61-64.
- [3] Song HY, Rothe M, Goeddel DV; Medline: 96270609 The tumor necrosis factor-inducible zinc finger protein A20 interacts with TRAF1/TRAF2 and inhibits NF-kappaB activation. Proc Natl Acad Sci U S A 1996;93:6721-6725.
 - [4] Opipari AW Jr, Boguski MS, Dixit VM; Medline: 90368626 The A20 cDNA induced by tumor necrosis factor alpha encodes a novel type of zinc finger protein." J Biol Chem 1990;265:14705-14708.

881. (zf-PARP)

Poly(ADP-ribose) polymerase zinc finger domain

Cross-reference(s) PS00347; PARP_ZN_FINGER_1 PS50064; PARP_ZN_FINGER_2

Poly(ADP-ribose) polymerase (EC 2.4.2.30) (PARP) [1,2] is a eukaryotic enzyme that catalyzes the covalent attachment of ADP-ribose units from NAD(+) to various nuclear

10

15

20

707

acceptor proteins. This post-translational modification of nuclear proteins is dependent on DNA. It appears to be involved in the regulation of various important cellular processes such as differentiation, proliferation and tumor transformation as well as in the regulation of the molecular events involved in the recovery of the cell from DNA damage. Structurally, PARP, about 1000 amino-acids residues long, consists of three distinct domains: an N-terminal zinc-dependent DNA-binding domain, a central automodification domain and a C-terminal NAD-binding domain. The DNA-binding region contains a pair of zinc finger domains which have been shown to bind DNA in a zinc-dependent manner. The zinc finger domains of PARP seem to bind specifically to single-stranded DNA. DNA ligase III [3] contains, in its N-terminal section, a single copy of a zinc finger highly similar to those of PARP.

Consensus pattern: C-[KR]-x-C-x(3)-I-x-K-x(3)-[RG]-x(16,18)-W-[FYH]-H-x(2)-C [The three C's and the H are zinc ligands] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROTNONE.

Note: This documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so.

- [1] Althaus F.R., Richter C.R. Mol. Biol. Biochem. Biophys. 37:1-126(1987).
- [2] de Murcia G., Menissier de Murcia J. Trends Biochem. Sci. 19:172-176(1994).
- [3] Wei Y.-F., Robins P., Carter K., Caldecott K., Pappin D.J.C., Yu G.-L., Wang R.-P., Shell B.K., Nash R.A., Schar P., Barnes D.E., Haseltine W.A., Lindahl T. Mol. Cell. Biol. 15:3206-3216(1995).
 - 882. Adenylylsulfate kinase (APS_kinase)
- Enzyme that catalyses the phosphorylation of adenylylsulfate to 3'-phosphoadenylylsulfate.

 This domain contains an ATP binding P-loop motif. Number of members: 34

[1] MacRae IJ, Rose AB, Segel IH; Medline: 99003196 Adenosine 5'-phosphosulfate kinase from Penicillium chrysogenum. site- directed mutagenesis at putative phosphoryl-accepting and ATP P-loop residues. J Biol Chem 1998;273:28583-28589.

5

10

15

883. DNA polymerase family B signature DNA_POLYMERASE_B (DNA_pol_B)

Replicative DNA polymerases (EC 2.7.7.7) are the key enzymes catalyzing the accurate replication of DNA. They require either a small RNA molecule or a protein as a primer for the de novo synthesis of a DNA chain. On the basis of sequence similarity, a number of DNA polymerases have been grouped [1 to 7] under the designation of DNA polymerase family B. These are:

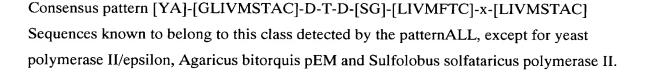
- Higher eukaryotes polymerases alpha.
- Higher eukaryotes polymerases delta.
- Yeast polymerase I/alpha (gene POL1), polymerase II/epsilon (gene POL2), polymerase III/delta (gene POL3) and polymerase REV3.
- Escherichia coli polymerase II (gene dinA or polB).
- Archaebacterial polymerases.
- Polymerases of viruses from the herpesviridae family.
- Polymerases from Adenoviruses.
- 20 Polymerases from Baculoviruses.
 - Polymerases from Chlorella viruses.
 - Polymerases from Poxviruses.
 - Bacteriophage T4 polymerase.
 - Podoviridae bacteriophages Phi-29, M2 and PZA polymerase.
- Tectiviridae bacteriophage PRD1 polymerase.
 - Polymerases encoded on mitochondrial linear DNA plasmids in various fungi and plants (Kluyveromyces lactis pGKL1 and pGKL2, Agaricus bitorquis pEM, Ascobolus immersus pAI2, Claviceps purpurea pCLK1, Neurospora Kalilo and Maranhar, maize S-1, etc).
- Six regions of similarity (numbered from I to VI) are found in all or a subset of the above polymerases. The most conserved region (I) includes a conserved tetrapeptide with two aspartate residues. Its function is not yet known. However, it has been suggested [3] that it may be involved in binding a magnesium ion. This conserved region was selected as a signature for this family of DNA polymerases.

20

25

30

5



[1] Jung G., Leavitt M.C., Hsieh J.-C., Ito J. Proc. Natl. Acad. Sci. U.S.A. 84:8287-8291(1987).

- [2] Bernad A., Zaballos A., Salas M., Blanco L. EMBO J. 6:4219-4225(1987).
- [3] Argos P. Nucleic Acids Res. 16:9909-9916(1988).
- 10 [4] Wang T.S.-F., Wong S.W., Korn D. FASEB J. 3:14-21(1989).
 - [5] Delarue M., Poch O., Todro N., Moras D., Argos P. Protein Eng. 3:461-467(1990).
 - [6] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).
 - [7] Braithwaite D.K., Ito J. Nucleic Acids Res. 21:787-802(1993).

884. DNA polymerase family X signature - DNA_POLYMERASE_X (DNA_polymeraseX)

DNA polymerases (EC 2.7.7.7) can be classified, on the basis of sequence similarity [1], into at least four different groups: A, B, C and X. DNA polymerases that belong to family X are listed below [2]:

- Vertebrate polymerase beta, involved in DNA repair.
- Yeast polymerase IV (POL4) [3], an enzyme with similar characteristics to that of the mammalian polymerase beta.
- Terminal deoxynucleotidyltransferase (TdT) (EC 2.7.7.31). TdT catalyzes the elongation of polydeoxynucleotide chains by terminal addition. One of the functions of this enzyme is the addition of nucleotides at the junction of rearranged Ig heavy chain and T cell receptor gene segments during the maturation of B and T cells.
 - African Swine Fever virus protein O174L [4].
 - Fission yeast hypothetical protein SpAC2F7.06c.

These enzymes are small (about 40 Kd) compared with other polymerases and their reaction mechanism operates via a distributive mode, i.e. they dissociate from the template-primer after addition of each nucleotide.

15

20

710

As a signature pattern for this family of DNA polymerases, a highly conserved region that contains a conserved arginine and two conserved aspartic acid residues were selected. The latter together with the arginine have been shown [5] to be involved in primer binding in polymerase beta.

Consensus pattern G-[SG]-[LFY]-x-R-[GE]-x(3)-[SGCL]-x-D-[LIVM]-D- [LIVMFY](3)-x(2)-[SAP] Sequences known to belong to this class detected by the patternALL.

- [1] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).
- 10 [2] Matsukage A., Nishikawa K., Ooi T., Seto Y., Yamaguchi M. J. Biol. Chem. 262:8960-8962(1987).
 - [3] Prasad R., Widen S.G., Singhal R.K., Watkins J., Prakash L., Wilson S.H. Nucleic Acids Res. 21:5301-5307(1993).
 - [4] Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C., Rodriguez J.F., Vinuela E. Virology 208:249-278(1995).
 - [5] Date T., Yamamoto S., Tanihara K., Nishimoto Y., Matsukage A. Biochemistry 30:5286-5292(1991).

885. DUF14 - Domain of unknown function

- This domain is found in glutamate synthase, tungsten formylmethanofuran dehydrogenase subunit c (FwdC) and molybdenum formylmethanofuran dehydrogenase subunit c (FmdC). It has no known function. Number of members: 52
- [1] Hochheimer A, Hedderich R, Thauer RK; Medline: 99035764. The formylmethanofuran dehydrogenase isoenzymes in Methanobacterium wolfei and Methanobacterium thermoautotrophicum: induction of the molybdenum isoenzyme by molybdate and constitutive synthesis of the tungsten isoenzyme." Arch Microbiol 1998;170:389-393.

886. DUF18-Domain of unknown function

- This domain of unknown function is found in several C. elegans proteins. The domain is 120 amino acids long and rich in cysteine residues. There are 16 conserved cysteine positions in the domain. Number of members: 34
 - 887. DUF27-Domain of unknown function

10

15

20

25

30

711

This domain is found in a number of otherwise unrelated proteins. This domain is found at the C-terminus of the macro-H2A histone protein Swiss:Q02874. This domain is found in the non-structural proteins of several types of ssRNA viruses such as NSP2 from alphaviruses Swiss:P03317. This domain is also found on its own in a family of proteins from bacteria Swiss:P75918, archaebacteria Swiss:O59182 and eukaryotes Swiss:Q17432, suggesting that it is involved in an important and ubiquitous cellular process. Number of members: 66

888. DUF37-Domain of unknown function

This domain is found in short (70 amino acid) hypothetical proteins from various bacteria. The domain contains three conserved cysteine residues. Swiss:Q44066 from Aeromonas hydrophila has been found to have hemolytic activity (unpublished). Number of members:

889. EGF-like domain signatures. (EGF-like)

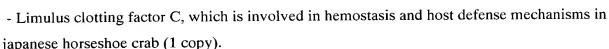
- A sequence of about thirty to forty amino-acid residues long found in the sequence of epidermal growth factor (EGF) has been shown [1 to 6] to be present, in a more or less conserved form, in a large number of other, mostly animal proteins. The proteins currently known to contain one or more copies of an EGF-like pattern are listed below.
 - Adipocyte differentiation inhibitor (gene PREF-1) from mouse (6 copies).
 - Agrin, a basal lamina protein that causes the aggregation of acetylcholine receptors on cultured muscle fibers (4 copies).
 - Amphiregulin, a growth factor (1 copy).
 - Betacellulin, a growth factor (1 copy).
 - Blastula proteins BP10 and Span from sea urchin which are thought to be involved in pattern formation (1 copy).
 - BM86, a glycoprotein antigen of cattle tick (7 copies).
 - Bone morphogenic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity (1-2 copies). Homologous proteins are found in sea urchin suBMP (1 copy) and in Drosophila the dorsal-ventral patterning protein tolloid (2 copies).
 - Caenorhabditis elegans developmental proteins lin-12 (13 copies) and glp-1 (10 copies).
 - Caenorhabditis elegans APX-1 protein, a patterning protein (4.5 copies).
 - Calcium-dependent serine proteinase (CASP) which degrades the extracellular matrix proteins type I and IV collagen and fibronectin (1 copy).

- Cartilage matrix protein CMP (1 copy).
- Cartilage oligomeric matrix protein COMP (4 copies).
- Cell surface antigen 114/A10 (3 copies).
- Cell surface glycoprotein complex transmembrane subunit ASGP-2 from rat (2 copies).
- 5 Coagulation associated proteins C, Z (2 copies) and S (4 copies).
 - Coagulation factors VII, IX, X and XII (2 copies).
 - Complement C1r components (1 copy).
 - Complement C1s components (1 copy).
 - Complement-activating component of Ra-reactive factor (RARF) (1 copy).
- Complement components C6, C7, C8 alpha and beta chains, and C9 (1 copy).
 - Crumbs, an epithelial development protein from Drosophila (29 copies).
 - Epidermal growth factor precursor (7-9 copies).
 - Exogastrula-inducing peptides A, C, D and X from sea urchin (1 copy).
 - Fat protein, a Drosophila cadherin-related tumor suppressor (5 copies).
- Fetal antigen 1, a probable neuroendocrine differentiation protein, which is derived from the delta-like protein (DLK) (6 copies).
 - Fibrillin 1 (47 copies) and fibrillin 2 (14 copies).
 - Fibropellins IA (21 copies), IB (13 copies), IC (8 copies), II (4 copies) and III (8 copies) from the apical lamina a component of the extracellular matrix of sea urchin.
- Fibulin-1 and -2, two extracellular matrix proteins (9-11 copies).
 - Giant-lens protein (protein Argos), which regulates cell determination and axon guidance in the Drosophila eye (1 copy).
 - Growth factor-related proteins from various poxviruses (1 copy).
 - Gurken protein, a Drosophila developmental protein (1 copy).
- Heparin-binding EGF-like growth factor (HB-EGF), transforming growth factor alpha (TGF-alpha), growth factors Lin-3 and Spitz (1 copy); the precursors are membrane proteins, the mature form is located extracellular.
 - Hepatocyte growth factor (HGF) activator (EC 3.4.21.-) (2 copies).
 - LDL and VLDL receptors, which bind and transport low-density lipoproteins and very low-density lipoproteins (3 copies).
 - LDL receptor-related protein (LRP), which may act as a receptor for endocytosis of extracellular ligands (22 copies).
 - Leucocyte antigen CD97 (3 copies), cell surface glycoprotein EMR1 (6 copies) and cell surface glycoprotein F4/80 (7 copies).

20

25

5



- Meprin A alpha subunit, a mammalian membrane-bound endopeptidase (1 copy).
- Milk fat globule-EGF factor 8 (MFG-E8) from mouse (2 copies).
- Neuregulin GGF-I and GGF-II, two human glial growth factors (1 copy).
 - Neurexins from mammals (3 copies).
 - Neurogenic proteins Notch, Xotch and the human homolog Tan-1 (36 copies), Delta (9 copies) and the similar differentiation proteins Lag-2 from Caenorhabditis elegans (2 copies), Serrate (14 copies) and Slit (7 copies) from Drosophila.
- Nidogen (also called entactin), a basement membrane protein from chordates (2-6 copies).
 - Ookinete surface proteins (24 Kd, 25 Kd, 28 Kd) from Plasmodium (4 copies).
 - Pancreatic secretory granule membrane major glycoprotein GP2 (1 copy).
 - Perforin, which lyses non-specifically a variety of target cells (1 copy).
 - Proteoglycans aggrecan (1 copy), versican (2 copies), perlecan (at least 2 copies), brevican (1 copy) and chondroitin sulfate proteoglycan (gene PG-M) (2 copies).
 - Prostaglandin G/H synthase 1 and 2 (EC 1.14.99.1) (1 copy), which is found in the endoplasmatic reticulum.
 - S1-5, a human extracellular protein whose ultimate activity is probably modulated by the environment (5 copies).
 - Schwannoma-derived growth factor (SDGF), an autocrine growth factor as well as a mitogen for different target cells (1 copy).
 - Selectins. Cell adhesion proteins such as ELAM-1 (E-selectin), GMP-140 (P-selectin), or the lymph-node homing receptor (L-selectin) (1 copy).
 - Serine/threonine-protein kinase homolog (gene Pro25) from Arabidopsis thaliana, which may be involved in assembly or regulation of light-harvesting chlorophyll A/B protein (2 copies).
 - Sperm-egg fusion proteins PH-30 alpha and beta from guinea pig (1 copy).
 - Stromal cell derived protein-1 (SCP-1) from mouse (6 copies).
 - TDGF-1, human teratocarcinoma-derived growth factor 1 (1 copy).
- Tenascin (or neuronectin), an extracellular matrix protein from mammals (14.5 copies), chicken (TEN-A) (13.5 copies) and the related proteins human tenascin-X (18 copies) and tenascin-like proteins TEN-A and TEN-M from Drosophila (8 copies).
 - Thrombomodulin (fetomodulin), which together with thrombin activates protein C (6 copies).

10

15

20

25

- Thrombospondin 1, 2 (3 copies), 3 and 4 (4 copies), adhesive glycoproteins that mediate cell-to-cell and cell-to-matrix interactions.
- Thyroid peroxidase 1 and 2 (EC 1.11.1.8) from human (1 copy).
- Transforming growth factor beta-1 binding protein (TGF-B1-BP) (16 or 18 copies).
- Tyrosine-protein kinase receptors Tek and Tie (EC 2.7.1.112) (3 copies).
 - Urokinase-type plasminogen activator (EC 3.4.21.73) (UPA) and tissue plasminogen activator (EC 3.4.21.68) (TPA) (1 copy).
 - Uromodulin (Tamm-horsfall urinary glycoprotein) (THP) (3 copies).
 - Vitamin K-dependent anticoagulants protein C (2 copies) and protein S (4 copies) and the similar protein Z, a single-chain plasma glycoprotein of unknown function (2 copies).
 - 63 Kd sperm flagellar membrane protein from sea urchin (3 copies).
 - 93 Kd protein (gene nel) from chicken (5 copies).
 - Hypothetical 337.6 Kd protein T20G5.3 from Caenorhabditis elegans (44 copies).
 - The functional significance of EGF domains in what appear to be unrelated proteins is not yet clear. However, a common feature is that these repeats are found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted (exception: prostaglandin G/H synthase). The EGF domain includes six cysteine residues which have been shown (in EGF) to be involved in disulfide bonds. The main structure is a two-stranded beta-sheet followed by a loop to a C-terminal short two-stranded sheet. Subdomains between the conserved cysteines strongly vary in length as shown in the following schematic representation of the EGF-like domain:

'C': conserved cysteine involved in a disulfide bond.

'G': often conserved glycine

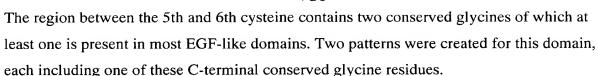
30 'a': often conserved aromatic amino acid

'*': position of both patterns.

'x': any residue

15

20



Consensus pattern: C-x-C-x(5)-G-x(2)-C [The 3 C's are involved in disulfide bonds]
Sequences known to belong to this class detected by the pattern A majority, but not those that have very long or very short regions between the last 3 conserved cysteines of their EGF-like domain(s). Other sequence(s) detected in SWISS-PROT87 proteins, of which 27 can be considered as possible candidates.

Consensus pattern: C-x-C-x(2)-[GP]-[FYW]-x(4,8)-C [The three C's are involved in disulfide bonds] Sequences known to belong to this class detected by the patternA majority, but not those that have very long or very short regions between the last 3 conserved cysteines of their EGF-like domain(s). Other sequence(s) detected in SWISS-PROT83 proteins, of which 49 can be considered as possible candidates. Note The beta chain of the integrin family of proteins contains 2 cysteine- rich repeats which were said to be dissimilar with the EGF pattern [7].

Note Laminin EGF-like repeats (see <PDOC00961>) are longer than the average EGF module and contain a further disulfide bond C-terminal of the EGF-like region. Perlecan and agrin contain both EGF-like domains and laminin-type EGF-like domains. Note the pattern do not detect all of the repeats of proteins with multiple EGF-like repeats. Note see <PDOC00913> for an entry describing specifically the subset of EGF- like domains that bind calcium.

25

30

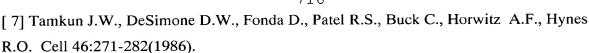
- [1] Davis C.G. New Biol. 2:410-419(1990).
- [2] Blomquist M.C., Hunt L.T., Barker W.C. Proc. Natl. Acad. Sci. U.S.A. 81:7363-7367(1984).
- [3] Barker W.C., Johnson G.C., Hunt L.T., George D.G. Protein Nucl. Acid Enz. 29:54-68(1986).
- [4] Doolittle R.F., Feng D.F., Johnson M.S. Nature 307:558-560(1984).
- [5] Appella E., Weber I.T., Blasi F. FEBS Lett. 231:1-4(1988).
- [6] Campbell I.D., Bork P. Curr. Opin. Struct. Biol. 3:385-392(1993).

15

20

25

30



5 890. Ham1 family (Ham1p_like)

This family consists of the HAM1 protein Swiss:P47119 and hypothetical archaeal bacterial and C. elegans proteins. HAM1 controls 6-N-hydroxylaminopurine (HAP) sensitivity and mutagenesis in S. cerevisiae Swiss:P47119 [1]. The HAM1 protein protects the cell from HAP, either on the level of deoxynucleoside triphosphate or the DNA level by a yet unidentified set of reactions [1]. Number of members: 19

[1] Noskov VN, Staak K, Shcherbakova PV, Kozmin SG, Negishi K, Ono BC, Hayatsu H, Pavlov YI; Medline: 96381244 HAM1, the gene controlling 6-N-hydroxylaminopurine sensitivity and mutagenesis in the yeast Saccharomyces cerevisiae." Yeast 1996;12:17-29.

891. (HCO3_cotransp)

Anion exchange is a cellular transport function which contributes to the regulation of cell pH and volume. Anion exchangers are a family of functionally related proteins that contributes to these properties by maintaining the intracellular level of the two principal anions: chloride and HCO3-. The best characterized anion exchanger is the band 3 protein [1], which is an erythrocyte anion exchange membrane glycoprotein. Band 3 is a protein of about 900 amino acids which consists of a cytoplasmic N-terminal domain of about 400 residues and an hydrophobic C-terminal section of about 500 residues that contains at least ten transmembrane regions. The cytoplasmic domain provides binding sites for cytoskeletal proteins, while the integral membrane domain is responsible for anion transport. Band 3 protein is specific to erythroid cells, at least two other proteins [2] structurally and functionally related to band 3, are found in nonerythroid tissues:

- AE2 (or B3 related protein; B3RP), a protein of 1200 residues, which seems to be present in a variety of cell types including lymphoid, kidney, and choroid plexus.
- AE3, a protein of 1200 residues, which is specific to neurons.

 Structurally AE2 and AE3 are very similar to band 3, the main difference being an extension of some 300 residues of the N-terminal domain in AE2 and AE3.

30

5

Two signature patterns were developed for these proteins. The first pattern is based on a conserved stretch of sequence that contains four clustered positive charged residues and which is located at the C-terminal extremity of the cytoplasmic domain, just before the first transmembrane segment from the integral domain. The second pattern is based on the perfectly conserved sequence of the fifth transmembrane segment; this segment contains a lysine, which is the covalent binding site for the isothiocyanate group of DIDS, an inhibitor of anion exchange.

Consensus pattern F-G-G-[LIVM](2)-[KR]-D-[LIVM]-[RK]-R-Y Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [FI]-L-I-S-L-I-F-I-Y-E-T-F-x-K-L Sequences known to belong to this class detected by the pattern ALL.

15 [1] Jay D., Cantley L. Annu. Rev. Biochem. 55:511-538(1986).

[2] Reithmeier R.A.F. Curr. Opin. Struct. Biol. 3:515-523(1993).

892. ATP phosphoribosyltransferase signature (HisG)

ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern a region located in the C-terminal part of this enzyme was selected.

Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM]

25 Sequences known to belong to this class detected by the pattern ALL.

893. HNH endonuclease (HNH)

Number of members: 56

- [1] Shub DA, Goodrich-Blair H, Eddy SR; Medline: 95117127 Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns."

 Trends Biochem Sci 1994;19:402-404.
 - [2] Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS; Medline: 98026854 Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases

10

15

20

25

30

and identification of an intein that encodes a site-specific endonuclease of the HNH family." Nucleic Acids Res 1997;25:4626-4638.

[3] Gorbalenya AE; Medline: 95004046 Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family." Protein Sci 1994;3:1117-1120.

894. NEUROHYPOPHYS HORM (hormone5)

Oxytocin (or ocytocin) and vasopressin [1] are small (nine amino acid residues), structurally and functionally related neurohypophysial peptide hormones. Oxytocin causes contraction of the smooth muscle of the uterus and of the mammary gland while vasopressin has a direct antidiuretic action on the kidney and also causes vasoconstriction of the peripheral vessels. Like the majority of active peptides, both hormones are synthesized as larger protein precursors that are enzymatically converted to their mature forms. Peptides belonging to this family are also found in birds, fish, reptiles and amphibians (mesotocin, isotocin, valitocin, glumitocin, aspargtocin, vasotocin, seritocin, asvatocin, phasvatocin), in worms (annetocin), octopi (cephalotocin), locust (locupressin or neuropeptide F1/F2) and in molluscs (conopressins G and S) [2]. The pattern developed to detect this category of peptides spans their entire sequence and includes four invariant amino acid residues.

Consensus pattern C-[LIFY](2)-x-N-[CS]-P-x-G [The two C's are linked by a disulfide bond]. Sequences known to belong to this class detected by the pattern ALL.

- [1] Acher R., Chauvet J. Biochimie 70:1197-1207(1988).
- [2] Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R. Int. J. Pept. Protein Res. 45:482-487(1995).

895. 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (HPPK)

All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Most microorganisms must synthesize folate de novo because they lack the active transport system of higher vertebrate cells which allows these organisms to use dietary folates. Enzymes involved in folate biosynthesis are therefore targets for a variety of antimicrobial agents such as trimethoprim or sulfonamides. 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (EC 2.7.6.3) (HPPK) catalyzes the attachment of pyrophosphate to 6-hydroxymethyl-7,8-dihydropterin to form 6-hydroxymethyl-7,8-dihydropteridine

10

15

20

25

30

pyrophosphate. This is the first step in a three-step pathway leading to 7,8-dihydrofolate. Bacterial HPPK (gene folk or sulD) [1] is a protein of 160 to 270 amino acids. In the lower eukaryote Pneumocystis carinii, HPPK is the central domain of a multifunctional folate synthesis enzyme (gene fas) [2]. As a signature for HPPK, a conserved region located in the central section of these enzymes was selected.

Consensus pattern [KRHD]-x-[GA]-[PSAE]-R-x(2)-D-[LIV]-D-[LIVM](2) Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Talarico T.L., Ray P.H., Dev I.K., Merrill B.M., Dallas W.S. J. Bacteriol. 174:5971-5977(1992).

[2] Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves C.J. Gene 112:213-218(1992).

896. Metalloenzyme superfamily (Metalloenzyme)

This family includes phosphopentomutase Swiss:P07651 and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, Swiss:P37689. This family is also related to alk_phosphatase [1]. The alignment contains the most conserved residues that are probably involved in metal binding and catalysis. Number of members: 34

[1] Galperin MY, Bairoch A, Koonin EV; Medline: 99180418 A superfamily of metalloenzymes unifies phosphopentomutase and cofactor- independent phosphoglycerate mutase with alkaline phosphatases and sulfatases." Protein Sci 1998;7:1829-1835.

897. Penicillin amidase (Penicil_amidase)

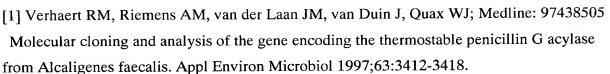
Penicillin amidase or penicillin acylase EC:3.5.1.11 catalyses the hydrolysis of benzylpenicillin to phenylacetic acid and 6-aminopenicillanic acid (6-APA) a key intermediate in the the synthesis of penicillins [1]. Also in the family is cephalosporin acylase Swiss:P07662 and Swiss:P29958 aculeacin A acylase which are involved in the synthesis of related peptide antibiotics. Number of members: 13

10

15

20

25



[2] Duggleby HJ, Tolley SP, Hill CP, Dodson EJ, Dodson G, Moody PC; Medline: 95115804 Penicillin acylase has a single-amino-acid catalytic centre." Nature 1995;373:264-268.

898. Phosphoribosyl-AMP cyclohydrolase (PRA-CH)

This enzyme catalyses the third step in the histidine biosynthetic pathway. It requires Zn ions for activity. Number of members: 13

[1] D'Ordine RL, Klem TJ, Davisson VJ; Medline: 99129952 N1-(5'-phosphoribosyl)adenosine-5'-monophosphate cyclohydrolase: purification and characterization of a unique metalloenzyme. Biochemistry 1999;38:1537-1546.

899. Phosphoribosyl-ATP pyrophosphohydrolase (PRA-PH)

This enzyme catalyses the second step in the histidine biosynthetic pathway. Number of members: 32

[1] Keesey JK Jr, Bigelis R, Fink GR; Medline: 79216449 The product of the his4 gene cluster in Saccharomyces cerevisiae. A trifunctional polypeptide." J Biol Chem 1979 Aug 10;254:7427-7433.

[2] Bruni CB, Carlomagno MS, Formisano S, Paolella G; Medline: 86310274 Primary and secondary structural homologies between the HIS4 gene product of Saccharomyces cerevisiae and the hisIE and hisD gene products of Escherichia coli and Salmonella typhimurium." Mol Gen Genet 1986;203:389-396.

30 900. Prokaryotic membrane lipoprotein lipid attachment site (PstS)

In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a

glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]):

- Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp).
- Escherichia coli lipoprotein-28 (gene nlpA).
- 5 Escherichia coli lipoprotein-34 (gene nlpB).
 - Escherichia coli lipoprotein nlpC.
 - Escherichia coli lipoprotein nlpD.
 - Escherichia coli osmotically inducible lipoprotein B (gene osmB).
 - Escherichia coli osmotically inducible lipoprotein E (gene osmE).
- Escherichia coli peptidoglycan-associated lipoprotein (gene pal).
 - Escherichia coli rare lipoproteins A and B (genes rplA and rplB).
 - Escherichia coli copper homeostasis protein cutF (or nlpE).
 - Escherichia coli plasmids traT proteins.
 - Escherichia coli Col plasmids lysis proteins.
- A number of Bacillus beta-lactamases.
 - Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA).
 - Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB).
 - Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7).
 - Chlamydia trachomatis outer membrane protein 3 (gene omp3).
- Fibrobacter succinogenes endoglucanase cel-3.
 - Haemophilus influenzae proteins Pal and Pcp.
 - Klebsiella pullulunase (gene pulA).
 - Klebsiella pullulunase secretion protein pulS.
 - Mycoplasma hyorhinis protein p37.
- Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlpABC).
 - Neisseria outer membrane protein H.8.
 - Pseudomonas aeruginosa lipopeptide (gene lppL).
 - Pseudomonas solanacearum endoglucanase egl.
 - Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC).
- 30 Rickettsia 17 Kd antigen.
 - Shigella flexneri invasion plasmid proteins mxiJ and mxiM.
 - Streptococcus pneumoniae oligopeptide transport protein A (gene amiA).
 - Treponema pallidium 34 Kd antigen.
 - Treponema pallidium membrane protein A (gene tmpA).

25

30

10

- 722
- Vibrio harveyi chitobiase (gene chb).
- Yersinia virulence plasmid protein yscJ.
- Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper-binding protein. This is the first archaebacterial protein known to be modified in such a fashion).
- From the precursor sequences of all these proteins, a consensus pattern was derived and a set of rules to identify this type of post-translational modification.

Consensus pattern {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTsome 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be.

- 15 [1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).
 - [2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).
 - [3] von Heijne G. Protein Eng. 2:531-534(1989).
 - [4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

901. Ribosome recycling factor (RRF)

The ribosome recycling factor (RRF / ribosome release factor) dissociates the ribosome from the mRNA after termination of translation, and is essential bacterial growth [1]. Thus ribosomes are "recycled" and ready for another round of protein synthesis. Number of members: 27

[1] Janosi L, Shimizu I, Kaji A; Medline: 94240115 Ribosome recycling factor (ribosome releasing factor) is essential for bacterial growth." Proc Natl Acad Sci U S A 1994;91:4249-4253.

902. S-layer homology(SLH)

20

25

5

723

S-layers are paracrystalline mono-layered assemblies of (glyco)proteins which coat the surface of bacteria [1]. Several S-layer proteins and some other cell wall proteins contain one or more copies of a domain of about 50-60 residues, which has been called SLH (for S-layer homology) [2]. There is strong evidence that this domain serves as an anchor to the peptidoglycan [3]. The SLH domain has been found in:

- S-layer glycoprotein of Acetogenium kivui (3 copies).
- S-layer 125 Kd protein of Bacillus sphaericus (3 copies).
- S-layer protein of Bacillus anthracis (3 copies).
- S-layer protein of Bacillus licheniformis (3 copies).
- S-layer protein (HWP) from Bacillus brevis strain HPD31 (3 copies).
 - Middle cell wall protein (MWP) from Bacillus brevis strain 47 (3 copies).
 - S-layer protein (p100) of Thermus thermophilus (1 copy).
 - Outer membrane protein Omp-alpha from Thermotoga maritima (1 copy).
 - Cellulosome anchoring protein (gene ancA), outer layer protein B (OlpB) and a further potential cell surface glycoprotein from Clostridium thermocellum (3 copies; the first copy is missing its N-terminal third which is appended to the end of the third copy; may have arisen by circular permutation).
 - Amylopullulanase (gene amyB) from Thermoanaerobacter thermosulfurogenes (3 copies)
 - Amylopullulanase (gene aapT) from Bacillus strain XAL-601 (3 copies).
 - Endoglucanase from Bacillus strain KSM-635 (3 copies).
 - Exoglucanase (gene xynX) from Clostridium thermocellum (3 copies).
 - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account).
 - Protein involved in butirosin production (ButB) from Bacillus circulans (2 incomplete copies; 3 copies if three frameshifts are taken into account).
 - Two hypothetical proteins from Synechocystis strain PCC 6803 (1 copy each).
 - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 copy; 3 copies if two frameshifts are taken into account).
- 30 SLH domains are found at the N- or C-termini of mature proteins. They occur in single copy followed by a predicted coiled coil domain, or in three contiguous copies. Structurally, the SLH domain is predicted to contain two alpha-helices flanking a beta strand. The SLH sequences are fairly divergent with an average identity of about 25%. It is however possible

20

25

to build a sequence pattern that starts at the second position of the domain and that spans 3/4 of its length.

Consensus pattern[LVFYT]-x-[DA]-x(2,5)-[DNGSATPHY]-[FYWPDA]-x(4)-[LIV]-x(2)[GTALV]-x(4,6)-[LIVFYC]-x(2)-G-x-[PGSTA]-x(2,3)-[MFYA]-x- [PGAV]-x(3,10)[LIVMA]-[STKR]-[RY]-x-[EQ]-x-[STALIVM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROTNONE.

- [1] Beveridge T.J. Curr. Opin. Struct. Biol. 4:204-212(1994).
- [2] Lupas A., Engelhardt H., Peters J., Santarius U., Volker S., Baumeister W. J. Bacteriol. 176:1224-1233(1994).
 - [3] Lemaire M., Ohayon H., Gounon P., Fujino T., Beguin P. J. Bacteriol. 177:2451-2459(1995).

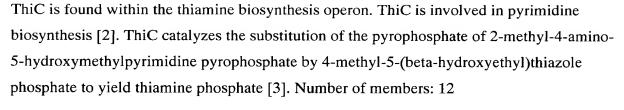
903. Queuine tRNA-ribosyltransferase (TGT)

This is a family of queuine tRNA-ribosyltransferases EC:2.4.2.29, also known as tRNA-guanine transglycosylase and guanine insertion enzyme. Queuine tRNA-ribosyltransferase modifies tRNAs for asparagine, aspartic acid, histidine and tyrosine with queuine. It catalyses the exchange of guanine-34 at the wobble position with 7-aminomethyl-7-deazaguanine, and the addition of a cyclopentenediol moiety to 7-aminomethyl-7-deazaguanine-34 tRNA; giving a hypermodified base queuine in the wobble position [1,2]. The aligned region contains a zinc binding motif C-x-C-x2-C-x29-H, and important tRNA and 7-aminomethyl-7-deazaguanine binding residues [1]. Number of members: 27

[1] Romier C, Reuter K, Suck D, Ficner R; Medline: 96256303 Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange." EMBO J 1996;15:2850-2857.

[2] Garcia GA, Koch KA, Chong S; Medline: 93287116 tRNA-guanine transglycosylase from Escherichia coli. Overexpression, purification and quaternary structure." J Mol Biol 1993;231:489-497.

904. ThiC Family (ThiC)



- [1] Vander Horn PB, Backstrom AD, Stewart V, Begley TP; Medline: 93163063 Structural genes for thiamine biosynthetic enzymes (thiCEFGH) in Escherichia coli K-12." J Bacteriol 1993;175:982-992.
- [2] Begley TP, Downs DM, Ealick SE, McLafferty FW, Van Loon AP, Taylor S, Campobasso N, Chiu HJ, Kinsland C, Reddick JJ, Xi J; Medline: 99311269 Thiamin biosynthesis in prokaryotes." Arch Microbiol 1999;171:293-300.
 - [3] Zhang Y, Taylor SV, Chiu HJ, Begley TP; Medline: 97284509 Characterization of the Bacillus subtilis thiC operon involved in thiamine biosynthesis." J Bacteriol 1997;179:3030-3035.

15

20

10

905. Putative tRNA binding domain (tRNA_bind)

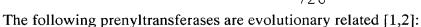
This domain is found in prokaryotic methionyl-tRNA synthetases, prokaryotic phenylalanyl tRNA synthetases the yeast GU4 nucleic-binding protein (G4p1 or p42, ARC1) [2], human tyrosyl-tRNA synthetase [1], and endothelial-monocyte activating polypeptide II. G4p1 binds specifically to tRNA form a complex with methionyl-tRNA synthetases [2]. In human tyrosyl-tRNA synthetase this domain may direct tRNA to the active site of the enzyme [2]. This domain may perform a common function in tRNA aminoacylation [1]. Number of members: 12

25

- [1] Kleeman TA, Wei D, Simpson KL, First EA; Medline: 97306356 Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine." J Biol Chem 1997;272:14420-14425.
- [2] Simos G, Segref A, Fasiolo F, Hellmuth K, Shevchenko A, Mann M, Hurt EC; Medline: 97050848 The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl-and glutamyl-tRNA synthetases." EMBO J 1996;15:5437-5448.

906. UbiA prenyltransferase family signature (UbiA)

10



- Bacterial 4-hydroxybenzoate octaprenyltransferase (gene ubiA).
- Yeast mitochondrial para-hydroxybenzoate--polyprenyltransferase (gene COQ2).
- Protoheme IX farnesyltransferase (heme O synthase) from yeast and mammals (gene
- 5 COX10) and from bacteria (genes cyoE or ctaB).

These proteins probably contain seven transmembrane segments. The best conserved region is located in a loop between the second and third of these segments and was used as a signature pattern.

Consensus pattern N-x(3)-[DE]-x(2)-[LIF]-D-x(2)-[VM]-x-R-[ST]-x(2)-R-x(4)-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROTNONE.

- 15 [1] Melzer M., Heide L. Biochim. Biophys. Acta 1212:93-102(1994).
 - [2] Mogi T., Saiki K., Anraku Y. Mol. Microbiol. 14:391-398(1994).
 - 907. Uncharacterized protein family UPF0044 signature (UPF0044)
- 20 The following uncharacterized proteins have been shown [1] to be highly similar:
 - Bacillus subtilis hypothetical protein yqeI.
 - Escherichia coli hypothetical protein yhbY and HI1333, the corresponding Haemophilus influenzae protein.
 - Methanococcus jannaschii hypothetical protein MJ0652.
- 25 These are small proteins of 10 to 15 Kd. They can be picked up in the database by the following pattern. This pattern is located in the N-terminal part of these proteins.

Consensus pattern L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)- [LIV]-[GA]-x(2)-G Sequences known to belong to this class detected by the patternALL.

908. ATP synthase (C/AC39) subunit (vATP-synt AC39)

This family includes the AC39 subunit from vacuolar ATP synthase Swiss:P32366 [1], and the C subunit from archaebacterial ATP synthase [2]. The family also includes subunit C

25

30

5

727

from the Sodium transporting ATP synthase from Enterococcus hirae Swiss:P43456 [3]. Number of members: 12

- [1] Bauerle C, Ho MN, Lindorfer MA, Stevens TH; Medline: 93286119 The Saccharomyces cerevisiae VMA6 gene encodes the 36-kDa subunit of the vacuolar H(+)-ATPase membrane sector." J Biol Chem 1993;268:12749-12757.
 - [2] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968 Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon Methanosarcina mazei Go1." J Biol Chem 1996;271:18843-18852.
- [3] Takase K, Kakinuma S, Yamato I, Konishi K, Igarashi K, Kakinuma Y; Medline: 94209269 Sequencing and characterization of the ntp gene cluster for vacuolar- type Na(+)-translocating ATPase of Enterococcus hirae." J Biol Chem 1994;269:11037-11044.
- 909. ATP synthase (E/31 kDa) subunit (vATP-synt_E)

 This family includes the vacuolar ATP synthase E subunit [1], as well as the archaebacterial ATP synthase E subunit [2]. Number of members: 24
 - [1] Foury F; Medline: 91009356 The 31-kDa polypeptide is an essential subunit of the vacuolar ATPase in Saccharomyces cerevisiae." J Biol Chem 1990;265:18554-18560.
 [2] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968
 Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon Methanosarcina mazei Go1." J Biol Chem 1996;271:18843-18852.

910. (WW)

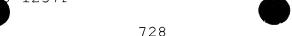
The WW domain [1-4,E1] (also known as rsp5 or WWP) has been originally discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown [5] to bind proteins with particular proline- motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro.

10

15

20

25



It is frequently associated with other domains typical for proteins in signal transduction processes.

Proteins containing the WW domain are listed below.

- Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C- terminal globular domain. Dystrophin form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats.
 - Utrophin, a dystrophin-like protein of unknown function.
- Vertebrate YAP protein is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains [6].
- Mouse NEDD-4 plays a role in the embryonic development and differentiation of the central nervous system. It contains 3 WW modules followed by a HECT domain. The human ortholog contains 4 WW domains, but the third WW domain is probably spliced resulting in an alternate NEDD-4 protein with only 3 WW modules [3].
- Yeast RSP5 is similar to NEDD-4 in its molecular organization. It contains an N-terminal C2 domain (see <PDOC00380>, followed by a histidine-rich region, 3 WW domains and a HECT domain.
- Rat FE65, a transcription-factor activator expressed preferentially in liver. The activator domain is located within the N-terminal 232 residues of FE65, which also contain the WW domain.
- Yeast ESS1/PTF1, a putative peptidyl prolyl cis-trans isomerase from family ppiC (see <PDOC00840>). A related protein, dodo (gene dod) exists in Drosophila and in mammals (gene PIN1).
- Tobacco DB10 protein. The WW domain is located N-terminal to the region with similarity to ATP-dependent RNA helicases.
 - IQGAP, a human GTPase activating protein acting on ras. It contains an N- terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain.

- Yeast pre-mRNA processing protein PRP40, Caenorhabditis elegans ZK1098.1 and fission yeast SpAC13C5.02 are related proteins with similarity to MYO2- type myosin, each containing two WW-domains at the N-terminus.
- Caenorhabditis elegans hypothetical protein C38D4.5, which contains one WW module, a PH domain (see <PDOC50003>) and a C-terminal phosphatidylinositol 3-kinase domain.
- Yeast hypothetical protein YFL010c.

 For the sensitive detection of WW domains, a profile was developed which spans the whole homology region as well as a pattern.
- Consensus pattern W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT8. Sequences known to belong to this class detected by the profileALL.
- 15 [1] Bork P., Sudol M. Trends Biochem. Sci. 19:531-533(1994).
 - [2] Andre B., Springael J.Y. Biochem. Biophys. Res. Commun. 205:1201-1205(1994).
 - [3] Hofmann K.O., Bucher P. FEBS Lett. 358:153-157(1995).
 - [4] Sudol M., Chen H.I., Bougeret C., Einbond A., Bork P. FEBS Lett. 369:67-71(1995).
 - [5] Chen H.I., Sudol M. Proc. Natl. Acad. Sci. U.S.A. 92:7819-7823(1995).
- [6] Sudol M., Bork P., Einbond A., Kastury K., Druck T., Negrini M., Huebner K., Lehman
 D. J. Biol. Chem. 270:14733-14741(1995).
 - 911. Xeroderma pigmentosum (XP) [1] (XPG_1)
- Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway:
 XP-A to XP-G. The defect in XP-G can be corrected by a 133 Kd nuclear protein called XPG (or XPGC) [2].

XPG belongs to a family of proteins [2,3,4,5,6] that are composed of two main subsets:

20

25

5

730

- Subset 1, to which belongs XPG, RAD2 from budding yeast and rad13 from fission yeast. RAD2 and XPG are single-stranded DNA endonucleases [7,8]. XPG makes the 3'incision in human DNA nucleotide excision repair [9].
- Subset 2, to which belongs mouse and human FEN-1, rad2 from fission yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific endonuclease.

In addition to the proteins listed in the above groups, this family also includes:

- Fission yeast exo1, a 5'->3' double-stranded DNA exonuclease that could act in a pathway that corrects mismatched base pairs.
- Yeast EXO1 (DHS1), a protein with probably the same function as exo1.
 - Yeast DIN7.

Sequence alignment of this family of proteins reveals that similarities are largely confined to two regions. The first is located at the N-terminal extremity (N-region) and corresponds to the first 95 to 105 amino acids. The second region is internal (I-region) and found towards the C-terminus; it spans about 140 residues and contains a highly conserved core of 27 amino acids that includes a conserved pentapeptide (E-A-[DE]-A-[QS]). It is possible that the conserved acidic residues are involved in the catalytic mechanism of DNA excision repair in XPG. The amino acids linking the N- and I-regions are not conserved; indeed, they are largely absent from proteins belonging to the second subset.

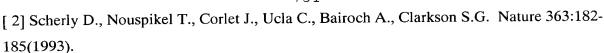
Two signature patterns were developed for these proteins. The first corresponds to the central part of the N-region, the second to part of the I-region and includes the putative catalytic core pentapeptide.

Consensus pattern [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

- Consensus pattern [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]- [QS]- [CLM] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.
 - [1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994).

20

30



- [3] Carr A.M., Sheldrick K.S., Murray J.M., Al-Harithy R., Watts F.Z., Lehmann A.R. Nucleic Acids Res. 21:1345-1349(1993).
- 5 [4] Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S., Lehmann A.R., Carr A.M., Watts F.Z. Mol. Cell. Biol. 14:4878-4888(1994).
 - [5] Harrington J.J., Lieber M.R. Genes Dev. 8:1344-1355(1994).
 - [6] Szankasi P., Smith G.R. Science 267:1166-1169(1995).
 - [7] Habraken Y., Sung P., Prakash L., Prakash S. Nature 366:365-368(1993).
- [8] O'Donovan A., Scherly D., Clarkson S.G., Wood R.D. J. Biol. Chem. 269:15965-15968(1994).
 - [9] O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D. Nature 371:432-435(1994).

912. 5-formyltetrahydrofolate cyclo-ligase (5-FTHF_cyc-lig)

5-formyltetrahydrofolate cyclo-ligase or methenyl-THF synthetase EC:6.3.3.2 catalyses the interchange of 5-formyltetrahydrofolate (5-FTHF) to 5-10-methenyltetrahydrofolate, this requires ATP and Mg2+ [1]. 5-FTHF is used in chemotherapy where it is clinically known as Leucovorin [2].

Number of members: 23

- [1] Dayan A, Bertrand R, Beauchemin M, Chahla D, Mamo A, Filion M, Skup D, Massie B,
 Jolivet J; Medline: 96096540 Cloning and characterization of the human 5,10-methenyltetrahydrofolate synthetase-encoding cDNA." Gene 1995;165:307-311.
 [2] Maras B, Stover P, Valiante S, Barra D, Schirch V; Medline: 94308074 Primary structure and tetrahydropteroylglutamate binding site of rabbit liver cytosolic 5,10-methenyltetrahydrofolate synthetase." J Biol Chem 1994;269:18429-18433.
 - 913. Cytosolic long-chain acyl-CoA thioester hydrolase (Acyl-CoA_hydro)

This family consist of various cytosolic long-chain acyl-CoA thioester hydrolases including human and rat [1,2]. The aligned region is repeated with in the sequence of human and rat

10

15

25

30

cytosolic long-chain acyl-CoA thioester hydrolases of this family. Long-chain acyl-CoA hydrolases hydrolyse palmitoyl-CoA to CoA and palmitate, they also catalyse the hydrolysis of other long chain fatty acyl-CoA thioesters. Long-chain acyl-CoA hydrolases are present in all living organisms and they may provide a mechanism for the control of lipid metabolism [1].

Number of members: 24

[1]Yamada J, Furihata T, Iida N, Watanabe T, Hosokawa M, Satoh T, Someya A, Nagaoka I, Suga T; Medline: 97236308 Molecular cloning and expression of cDNAs encoding rat brain and liver cytosolic long-chain acyl-CoA hydrolases." Biochem Biophys Res Commun 1997;232:198-203.

[2] Broustas CG, Larkins LK, Uhler MD, Hajra AK; Medline: 96209964 Molecular cloning and expression of cDNA encoding rat brain cytosolic acyl-coenzyme A thioester hydrolase." J Biol Chem 1996;271:10470-10476.

914. Agglutinin

Lectin (probable mannose binding)

Members of this family are plant lectins. Many if not all are mannose specific.

20 Number of members: 87

[1] Wright CS, Hester G; Medline: 97094989 The 2.0 A structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: evidence for two unique binding modes." Structure 1996;4:1339-1352.

915. (ANF RECEPTORS)

Natriuretic peptides are hormones involved in the regulation of fluid and electrolyte homeostasis. These hormones stimulate the intracellular production of cyclic GMP as a second messenger.

Currently, three types of natriuretic peptide receptors are known [1,2]. Two express guanylate cyclase activity: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which seems to be stimulated more effectively by brain natriuretic

10

15

733

peptide (BNP) than by ANP. The third receptor (ANP-C) is probably responsible for the clearance of ANP from the circulation and does not play a role in signal transduction.

GC-A and GC-B are plasma membrane-bound proteins that share the following topology: an N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain followed by a large cytoplasmic C- terminal region that can be subdivided into two domains: a protein kinase-like domain (see <PDOC00100>) that appears important for proper signalling and a guanylate cyclase catalytic domain (see <PDOC00425>). The topology of ANP-C is different: like GC-A and -B it possesses an extracellular ligand-binding region and a transmembrane domain, but its cytoplasmic domain is very short.

A pattern was developed from the ligand-binding region of natriuretic peptide receptors based on a highly conserved region located in the N-terminal part of the domain.

Consensus patternG-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

- [1] Garbers D.L. New Biol. 2:499-504(1990).
- 20 [2] Schulz S., Chinkers M., Garbers D.L. FASEB J. 2:2026-2035(1989).

916. (Apocytochrome)

Cytochrome c family heme-binding site signature

In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c.

Consensus patternC-{CPWHF}-{CPWR}-C-H-{CFYW} Sequences known to belong to this class detected by the patternALL, except for four cytochrome c's which lack the first thioether bond. Other sequence(s) detected in SWISS-PROT454.

20

25

30

Note: some cytochrome c's have more than a single bound heme groupc4 has 2, c7 has 3, c3 has 4, the reaction center has 4, and cc3/Hmc has 16!

5 [1] Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).

917. ATP-synt_A-c. ATP synthase Alpha chain, C terminal

[1] Medline: 94344236. Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. Abrahams JP, Leslie AG, Lutter R, Walker JE; Nature 1994;370:621-628.

10 Number of members: 125

918. (Basic)

Myc-type, 'helix-loop-helix' dimerization domain signature

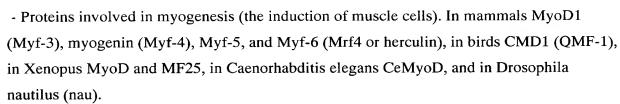
HELIX_LOOP_HELIX

A number of eukaryotic proteins, which probably are sequence specific DNA- binding proteins that act as transcription factors, share a conserved domain of 40 to 50 amino acid residues. It has been proposed [1] that this domain is formed of two amphipathic helices joined by a variable length linker region that could form a loop. This 'helix-loop-helix' (HLH) domain mediates protein dimerization and has been found in the proteins listed below [2,3,E1,E2]. Most of these proteins have an extra basic region of about 15 amino acid residues that is adjacent to the HLH domain and specifically binds to DNA. They are refered as basic helix-loop-helix proteins (bHLH), and are classified in two groups: class A (ubiquitous) and class B (tissue-specific). Members of the bHLH family bind variations on the core sequence 'CANNTG', also refered to as the E-box motif. The homo- or heterodimerization mediated by the HLH domain is independent of, but necessary for DNA binding, as two basic regions are required for DNA binding activity. The HLH proteins lacking the basic domain (Emc, Id) function as negative regulators since they form heterodimers, but fail to bind DNA. The hairy-related proteins (hairy, E(spl), deadpan) also repress transcription although they can bind DNA. The proteins of this subfamily act together with co-repressor proteins, like groucho, through their C-terminal motif WRPW.

- The myc family of cellular oncogenes [4], which is currently known to contain four members: c-myc [E3], N-myc, L-myc, and B-myc. The myc genes are thought to play a role in cellular differentiation and proliferation.

20

25



- Vertebrate proteins that bind specific DNA sequences ('E boxes') in various immunoglobulin chains enhancers: E2A or ITF-1 (E12/pan-2 and E47/pan-1), ITF-2 (tcf4), TFE3, and TFEB.
 - Vertebrate neurogenic differentiation factor 1 that acts as differentiation factor during neurogenesis.
- Vertebrate MAX protein, a transcription regulator that forms a sequence- specific DNAbinding protein complex with myc or mad.
 - Vertebrate Max Interacting Protein 1 (MXI1 protein) which acts as a transcriptional repressor and may antagonize myc transcriptional activity by competing for max.
 - Proteins of the bHLH/PAS superfamily which are transcriptional activators. In mammals,
 - AH receptor nuclear translocator (ARNT), single-minded homologs (SIM1 and SIM2), hypoxia-inducible factor 1 alpha (HIF1A), AH receptor (AHR), neuronal pas domain proteins (NPAS1 and NPAS2), endothelial pas domain protein 1 (EPAS1), mouse ARNT2, and human BMAL1. In drosophila, single-minded (SIM), AH receptor nuclear translocator (ARNT), trachealess protein (TRH), and similar protein (SIMA).
 - Mammalian transcription factors HES, which repress transcription by acting on two types of DNA sequences, the E box and the N box.
 - Mammalian MAD protein (max dimerizer) which acts as transcriptional repressor and may antagonize myc transcriptional activity by competing for max.
 - Mammalian Upstream Stimulatory Factor 1 and 2 (USF1 and USF2), which bind to a symmetrical DNA sequence that is found in a variety of viral and cellular promoters.
 - Human lyl-1 protein; which is involved, by chromosomal translocation, in T- cell leukemia.
 - Human transcription factor AP-4.
 - Mouse helix-loop-helix proteins MATH-1 and MATH-2 which activate E box- dependent transcription in collaboration with E47.
- Mammalian stem cell protein (SCL) (also known as tal1), a protein which may play an important role in hemopoietic differentiation. SCL is involved, by chromosomal translocation, in stem-cell leukemia.

20

30

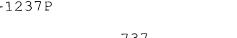
5



- Mammalian proteins Id1 to Id4 [5]. Id (inhibitor of DNA binding) proteins lack a basic DNA-binding domain but are able to form heterodimers with other HLH proteins, thereby inhibiting binding to DNA.
- Drosophila extra-macrochaetae (emc) protein, which participates in sensory organ patterning by antagonizing the neurogenic activity of the achaete- scute complex. Emc is the homolog of mammalian Id proteins.
- Human Sterol Regulatory Element Binding Protein 1 (SREBP-1), a transcriptional activator that binds to the sterol regulatory element 1 (SRE-1) found in the flanking region of the LDLR gene and in other genes.
- Drosophila achaete-scute (AS-C) complex proteins T3 (l'sc), T4 (scute), T5 (achaete) and T8 (asense). The AS-C proteins are involved in the determination of the neuronal precursors in the peripheral nervous system and the central nervous system.
 - Mammalian homologs of achaete-scute proteins, the MASH-1 and MASH-2 proteins.
 - Drosophila atonal protein (ato) which is involved in neurogenesis.
 - Drosophila daughterless (da) protein, which is essential for neurogenesis and sexdetermination.
 - Drosophila deadpan (dpn), a hairy-like protein involved in the functional differentiation of neurons.
 - Drosophila delilah (dei) protein, which is plays an important role in the differentiation of epidermal cells into muscle.
 - Drosophila hairy (h) protein, a transcriptional repressor which regulates the embryonic segmentation and adult bristle patterning.
 - Drosophila enhancer of split proteins E(spl), that are hairy-like proteins active during neurogenesis. also act as transcriptional repressors.
- Drosophila twist (twi) protein, which is involved in the establishment of germ layers in embryos.
 - Maize anthocyanin regulatory proteins R-S and LC.
 - Yeast centromere-binding protein 1 (CPF1 or CBF1). This protein is involved in chromosomal segregation. It binds to a highly conserved DNA sequence, found in centromers and in several promoters.
 - Yeast INO2 and INO4 proteins.
 - Yeast phosphate system positive regulatory protein PHO4 which interacts with the upstream activating sequence of several acid phosphatase genes.
 - Yeast serine-rich protein TYE7 that is required for ty-mediated ADH2 expression.

25

30



- Neurospora crassa nuc-1, a protein that activates the transcription of structural genes for phosphorus acquisition.
- Fission yeast protein esc1 which is involved in the sexual differentiation process.
- The schematic representation of the helix-loop-helix domain is shown here: 5 helix 1 Loop Amphipathic helix 2

The signature pattern that had been developed to detect this domain spans completely the second amphipathic helix. 10

Consensus pattern[DENSTAP]-[KR]-[LIVMAGSNT]-{FYWCPHKR}-[LIVMT]-[LIVM]x(2)-[STAV]-[LIVMSTACKR]-x-[VMFYH]-[LIVMTA]- $\{P\}$ - $\{P\}$ - [LIVMRKHQ] Sequences known to belong to this class detected by the pattern the majority but far from all. Other sequence(s) detected in SWISS-PROT135.

- [1] Murre C., McCaw P.S., Baltimore D. Cell 56:777-783(1989).
- [2] Garrel J., Campuzano S. BioEssays 13:493-498(1991).
- [3] Kato G.J., Dang C.V. FASEB J. 6:3065-3072(1992).
- [4] Krause M., Fire A., Harrison S.W., Priess J., Weintraub H. Cell 63:907-919(1990). 2.0
 - [5] Riechmann V., van Cruechten I., Sablitzky F. Nucleic Acids Res. 22:749-755(1994).

919. (Beta-lactamase)

Beta-lactamases classes -A, -C, and -D active site

Beta-lactamases (EC 3.5.2.6) [1,2] are enzymes which catalyze the hydrolysis of an amide bond in the beta-lactam ring of antibiotics belonging to the penicillin/cephalosporin family. Four kinds of beta-lactamase have been identified [3]. Class-B enzymes are zinc containing proteins whilst class -A, C and D enzymes are serine hydrolases. The three classes of serine beta-

lactamases are evolutionary related and belong to a superfamily [4] that also includes DDpeptidases and a variety of other penicillin-binding proteins (PBP's). All these proteins contain a Ser-x-x-Lys motif, where the serine is the active site residue. Although clearly homologous, the sequences of the three classes of serine beta-lactamases exhibit a large

catalytic serine.

each of the three classes.

5

10

15

20

25

30



Since a pattern detecting all serine beta-lactamases would also pick up many unrelated

sequences, it was decided to provide specific patterns, centered on the active site serine, for

degree of variability and only a small number of residues are conserved in addition to the

Consensus pattern [FY]-x-[LIVMFY]-x-S-[TV]-x-K-x(4)-[AGLM]-x(2)-[LC] [S is the active site residue] Sequences known to belong to this class detected by the patternALL class-A

Consensus pattern F-E-[LIVM]-G-S-[LIVMG]-[SA]-K [The first S is the active site residue] Sequences known to belong to this class detected by the patternALL class-C beta-lactamases. Other sequence(s) detected in SWISS-PROTNONE.

Consensus pattern [PA]-x-S-[ST]-F-K-[LIV]-[PAL]-x-[STA]-[LI] [S is the active site residue] Sequences known to belong to this class detected by the patternALL class-D betalactamases. Other sequence(s) detected in SWISS-PROTNONE.

- [1] Ambler R.P. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 289:321-331(1980).
 - [2] Pastor N., Pinero D., Valdes A.M., Soberon X. Mol. Microbiol. 4:1957-1965(1990).
 - [3] Bush K. Antimicrob. Agents Chemother. 33:259-263(1989).

beta-lactamases. Other sequence(s) detected in SWISS-PROT7.

[4] Joris B., Ghuysen J.-M., Dive G., Renard A., Dideberg O., Charlier P., Frere J.M., Kelly J.A., Boyington J.C., Moews P.C., Knox J.R. Biochem. J. 250:313-324(1988).

920. Biotin protein ligase (BPL)

Biotin is covalently attached at the active site of certain enzymes that transfer carbon dioxide from bicarbonate to organic acids to form cellular metabolites. Biotin protein ligase (BPL) is the enzyme responsible for attaching biotin to a specific lysine at the active site of biotin enzymes. Each organism probably has only one BPL. Biotin attachment is a two step reaction that results in the formation of an amide linkage between the carboxyl group of biotin and the epsilon-amino group of the modified lysine [2].

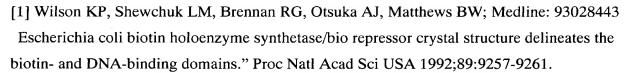
Number of members: 26

10

15

20

30



[2] Chapman-Smith A, Cronan JE Jr; Medline: 10470036 The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity." Trends Biochem Sci 1999;24:359-363.

921. (BRCA2_repeat)

The alignment covers only the most conserved region of the repeat. Respiratory-chain NADH dehydrogenase 30 Kd subunit signature

[1] Bork P, Blomberg N, Nilges M; Medline: 96241568 Internal repeats in the BRCA2 protein sequence." Nat Genet 1996;13:22-23.

Number of members: 63

922. (C6)

This domain of unknown function is found in the C. elegans protein Swiss:Q19522. It is presumed to be an extracellular domain. The C6 domain contains six conserved cysteine residues in most copies of the domain. However some copies of the domain are missing cysteine residues 1 and 3 suggesting that these form a disulphide bridge.

25 Number of members: 23

923. Cadherin cytoplasmic region (Cadherin_C_term)

Cadherins are vital in cell-cell adhesion during tissue differentiation. Cadherins are linked to the cytoskeleton by catenins. Catenins bind to the cytoplasmic tail of the cadherin. Cadherins cluster to form foci of homophilic binding units. A key determinant to the strength of the binding that it is mediated by cadherins is the juxtamembrane region of the cadherin. This region induces clustering and also binds to the protein p120ctn [1].

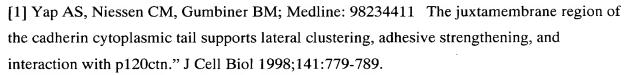
Number of members: 59

15

20

25

30



- [2] Barth AI, Nathke IS, Nelson WJ; Medline: 97471931 Cadherins, catenins and APC protein: interplay between cytoskeletal complexes and signaling pathways." Curr Opin Cell Biol 1997;9:683-690.
 - [3] Braga VM, Machesky LM, Hall A, Hotchin NA; Medline: 97327766 The small GTPases Rho and Rac are required for the establishment of cadherin-dependent cell-cell contacts." J Cell Biol 1997;137:1421-1431.

924. Clathrin propeller repeat (Clathrin_propel)

Clathrin is the scaffold protein of the basket-like coat that surrounds coated vesicles. The soluble assembly unit, a triskelion, contains three heavy chains and three light chains in an extended three-legged structure. Each leg contains one heavy and one light chain. The N-terminus of the heavy chain is known as the globular domain, and is composed of seven repeats which form a beta propeller [1].

Number of members: 61

[1] ter Haar E, Musacchio A, Harrison SC, Kirchhausen T; Medline: 99043510 Atomic structure of clathrin: a beta propeller terminal domain joins an alpha zigzag linker." Cell. 1998;95:563-573.

925. Respiratory-chain NADH dehydrogenase 30 Kd subunit signature (complex1_30Kd)

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 30 Kd (in mammals) which has been found to be:

- Nuclear encoded, as a precursor form with a transit peptide in mammals, and in Neurospora crassa.



- Mitochondrial encoded in Paramecium (protein P1), and in the slime mold Dictyostelium discoideum (ORF 209).
- Chloroplast encoded in various higher plants (ORF 159). It is also present in bacteria:
- In the cyanobacteria Synechocystis strain PCC 6803 (gene ndhJ).
- Subunit C of Escherichia coli NADH-ubiquinone oxidoreductase (gene nuoC).
- Subunit NQO5 of Paracoccus denitrificans NADH-ubiquinone oxidoreductase.

This protein, in its mature form, consists of from 157 to 266 amino acid residues. The best conserved region is located in the C-terminal section and can be used as a signature pattern.

10

5

Consensus pattern E-R-E-x(2)-[DE]-[LIVMFY](2)-x(6)-[HK]-x(3)-[KRP]-x-[LIVM]-[LIVMYS] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987). 15

- [2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).
- 926. Respiratory-chain NADH dehydrogenase 49 Kd subunit signature (complex1_49Kd)
- Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or 20 NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 49 Kd (in mammals), which is the third largest subunit of complex I and is a component of the 25 iron-sulfur (IP) fragment of the enzyme. It seems to bind a 4Fe-4S iron-sulfur cluster. The 49 Kd subunit has been found to be:
 - Nuclear encoded, as a precursor form with a transit peptide in mammals, and in Neurospora crassa.
- Mitochondrial encoded in protozoan such as Paramecium (ORF 400), Leishmania and 30 Trypanosoma (MURF 3).
 - Chloroplast encoded in various higher plants (ORF 392).

The 49 Kd subunit is highly similar to [3,4]:

- Subunit D of Escherichia coli NADH-ubiquinone oxidoreductase (gene nuoD).

20

25

30

5



- Subunit NQO4 of Paracoccus denitrificans NADH-ubiquinone oxidoreductase.
- Subunit 5 of Escherichia coli formate hydrogenlyase (gene hycE).
- Subunit G of Escherichia coli hydrogenase-4 (gene hyfG).

A highly conserved region was selected as signature pattern, located in the N-terminal section of this subunit.

Consensus pattern [LIVMH]-H-[RT]-[GA]-x-E-K-[LIVMTN]-x-E-x-[KRQ] Sequences known to belong to this class detected by the patternALL.

- 10 [1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).
 - [2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).
 - [3] Fearnley I.M., Walker J.E. Biochim. Biophys. Acta 1140:105-134(1992).
 - [4] Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).

927. (COX2)

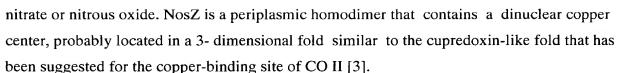
Cytochrome c oxidase (EC 1.9.3.1) [1,2] is an oligomeric enzymatic complex which is a component of the respiratory chain and is involved in the transfer of electrons from cytochrome c to oxygen. In eukaryotes this enzyme complex is located in the mitochondrial inner membrane; in aerobic prokaryotes it is found in the plasma membrane. The enzyme complex consists of 3-4 subunits (prokaryotes) to up to 13 polypeptides (mammals).

Subunit 2 (CO II) transfers the electrons from cytochrome c to the catalytic subunit 1. It contains two adjacent transmembrane regions in its N-terminus and the major part of the protein is exposed to the periplasmic or to the mitochondrial intermembrane space, respectively. CO II provides the substrate- binding site and contains a copper center called Cu(A), probably the primary acceptor in cytochrome c oxidase. An exception is the corresponding subunit of the cbb3-type oxidase which lacks the copper A redox-center. Several bacterial CO II have a C-terminal extension that contains a covalently bound heme c.

It has been shown [3,4] that nitrous oxide reductase (EC 1.7.99.6) (gene nosZ) of Pseudomonas has sequence similarity in its C-terminus to CO II. This enzyme is part of the bacterial respiratory system which is activated under anaerobic conditions in the presence of

25

30



- The dinuclear purple copper center is formed by 2 histidines and 2 cysteines [5]. This region was used as a signature pattern. The conserved valine and the conserved methionine are said to be involved in stabilizing the copper-binding fold by interacting with each other.
- Consensus pattern V-x-H-x(33,40)-C-x(3)-C-x(3)-H-x(2)-M [The two C's and two H's are copper ligands] Sequences known to belong to this class detected by the patternALL, except for Paramecium primaurelia as well as in some plants where the pattern ends with Thr; an RNA editing event at this position could change this Thr to Met.

Note: cytochrome cbb(3) subunit 2 does not belong to this family.

- [1] Capaldi R.A., Malatesta F., Darley-Usmar V.M. Biochim. Biophys. Acta 726:135-148(1983).
- [2] Garcia-Horsman J.A., Barquera B., Rumbley J., Ma J., Gennis R.B. J. Bacteriol. 176:5587-5600(1994).
- [3] van der Oost J., Lappalainen P., Musacchio A., Warne A., Lemieux L., Rumbley J., Gennis R.B., Aasa R., Pascher T., Malmstrom B.G., Saraste M. EMBO J. 11:3209-3217(1992).
 - [4] Zumft W.G., Dreutsch A., Loechelt S., Cuypers H., Friedrich B., Schneider B. Eur. J. Biochem. 208:31-40(1992).
 - 928. Cytochrome C assembly protein (CytC_asm)

This family consists of various proteins involved in cytochrome c assembly from mitochondria and bacteria; CycK from Rhizobium[3], CcmC from E. coli and Paracoccus denitrificans [2,1] and orf240 from wheat mitochondria [4]. The members of this family are probably integral membrane proteins with six predicted transmembrane helices. It has been proposed that members of this family comprise a membrane component of an ABC (ATP binding cassette) transporter complex. It is also proposed that this transporter is necessary for transport of some component needed for cytochrome c assembly. One member CycK

Number of members:

10

15

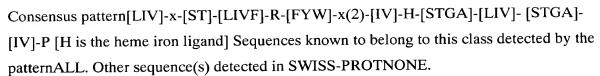
20

contains a putative heme-binding motif [3], orf240 also contains a putative heme-binding motif and is a proposed ABC transporter with c-type heme as its proposed substrate [4]. However it seems unlikely that all members of this family transport heme nor c-type apocytochromes because CcmC in the putative CcmABC transporter transports neither [1].

- [1] Page D, Pearce DA, Norris HA, Ferguson SJ; Medline: 97195802 The Paracoccus denitrificans ccmA, B and C genes: cloning and sequencing, and analysis of the potential of their products to form a haem or apo-c-type cytochrome transporter. MICROBIOLOGY 1997;143:563-576.
- [2] Thoeny-meyer L, Fischer F, Kunzler P, Ritz D, Hennecke H; Medline: 95362656 Escherichia coli genes required for cytochrome c maturation." J. BACTERIOL 1995;177:4321-4326.
- [3] Delgado MJ, Yeoman KH, Wu G, Vargas C, Davies A, Poole RK, Johnston AWB, Downie JA; Medline: 95394794 Characterization of the cycHJKL genes involved in cytochrome c biogenesis and symbiotic nitrogen fixation in Rhizobium leguminosarum." J. BACTERIOL 1995;177:4927-4934.
- [4] Bonnard G, Grienenberger JM; Medline: 95124303 A gene proposed to encode a transmembrane domain of an ABC transporter is expressed in wheat mitochondria." MOL. GEN. GENET 1995;246:91-99.
- 929. Cytochrome b559 subunits heme-binding site signature (cytochr_b559)
- Cytochrome b559 [1] is an essential component of photosystem II complex from oxygenic photosynthetic organisms. It is an integral thylakoid membrane protein composed of two subunits, alpha (gene psbE) and beta (gene psbF), each of which contains a histidine residue located in a transmembrane region. The two histidines coordinate the heme iron of cytochrome b559.
- The region around the heme-binding residue of both subunits is very similar and can be used as a signature pattern.

20

25



5 [1] Pakrasi H.B., de Ciechi P., Whitmarsh J. EMBO J. 10:1619-1627(1991).

930. Cytochrome b/b6 signatures (Cytochrome_b)

In the mitochondrion of eukaryotes and in aerobic prokaryotes, cytochrome b is a component of respiratory chain complex III (EC 1.10.2.2) - also known as the bc1 complex or ubiquinol-cytochrome c reductase. In plant chloroplasts and cyanobacteria, there is a analogous protein, cytochrome b6, a component of the plastoquinone-plastocyanin reductase (EC 1.10.99.1), also known as the b6f complex.

Cytochrome b/b6 [1,2] is an integral membrane protein of approximately 400 amino acid residues that probably has 8 transmembrane segments. In plants and cyanobacteria, cytochrome b6 consists of two subunits encoded by the petB and petD genes. The sequence of petB is colinear with the N-terminal part of mitochondrial cytochrome b, while petD corresponds to the C-terminal part. Cytochrome b/b6 non-covalently binds two heme groups, known as b562 and b566. Four conserved histidine residues are postulated to be the ligands of the iron atoms of these two heme groups.

Apart from regions around some of the histidine heme ligands, there are a few conserved regions in the sequence of b/b6. The best conserved of these regions includes an invariant P-E-W triplet which lies in the loop that separates the fifth and sixth transmembrane segments. It seems to be important for electron transfer at the ubiquinone redox site - called Qz or Qo (where o stands for outside) - located on the outer side of the membrane.

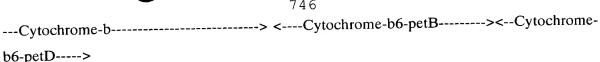
30 A schematic representation of the structure of cytochrome b/b6 is shown below.

10

15

25

30



Two signature patterns were developed for cytochrome b/b6. The first includes the first conserved histidine of b/b6, which is a heme b562 ligand; the second includes the conserved PEW triplet.

Consensus pattern [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H [H is a heme b562 ligand] Sequences known to belong to this class detected by the patternALL, except for 5 sequences.

Consensus pattern P-[DE]-W-[FY]-[LFY](2) Sequences known to belong to this class detected by the patternALL, except for Odocoileus hemionus (mule deer) and Paramecium tetraurelia cytochrome b.

- [1] Howell N. J. Mol. Evol. 29:157-169(1989).
- [2] Esposti M.D., de Vries S., Crimi M., Ghelli A., Patarnello T., Meyer A. Biochim. Biophys. Acta 1143:243-271(1993).
- 931. Phorbol esters / diacylglycerol binding domain (DAG_PE-bind) 20

Diacylglycerol (DAG) is an important second messenger. Phorbol esters (PE) are analogues of DAG and potent tumor promoters that cause a variety of physiological changes when administered to both cells and tissues. DAG activates a family of serine/threonine protein kinases, collectively known as protein kinase C (PKC) [1]. Phorbol esters can directly stimulate PKC. The N- terminal region of PKC, known as C1, has been shown [2] to bind PE and DAG in a phospholipid and zinc-dependent fashion. The C1 region contains one or two copies (depending on the isozyme of PKC) of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. Such a domain has also been found in the following proteins:

- Diacylglycerol kinase (EC 2.7.1.107) (DGK) [3], the enzyme that converts DAG into phosphatidate. It contains two copies of the DAG/PE-binding domain in its N-terminal section. At least five different forms of DGK are known in mammals.

10

15

20



- N-chimaerin. A brain specific protein which shows sequence similarities with the BCR protein at its C-terminal part and contains a single copy of the DAG/PE-binding domain at its N-terminal part. It has been shown [4,5] to be able to bind phorbol esters.
- The raf/mil family of serine/threonine protein kinases. These protein kinases contain a single N-terminal copy of the DAG/PE-binding domain.
- The unc-13 protein from Caenorhabditis elegans. Its function is not known but it contains a copy of the DAG/PE-binding domain in its central section and has been shown to bind specifically to a phorbol ester in the presence of calcium [6].
- The vav oncogene. Vav was generated by a genetic rearrangement during gene transfer assays. Its expression seems to be restricted to cells of hematopoeitic origin. Vav seems [5,7] to contain a DAG/PE-binding domain in the central part of the protein.
 - The Drosophila GTPase activating protein rotund.

The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain. A signature pattern was developed that spans completely the DAG/PE domain.

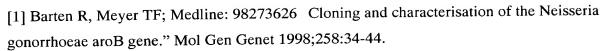
Consensus pattern H-x-[LIVMFYW]-x(8,11)-C-x(2)-C-x(3)-[LIVMFC]-x(5,10)- C-x(2)-C-x(4)-[HD]-x(2)-C-x(5,9)-C [All the C and H are involved in binding Zinc] Sequences known to belong to this class detected by the pattern ALL, except a few DGK's.

- [1] Azzi A., Boscoboinik D., Hensey C. Eur. J. Biochem. 208:547-557(1992).
- [2] Ono Y., Fujii T., Igarashi K., Kuno T., Tanaka C, Kikkawa U., Nishizuka Y. Proc. Natl. Acad. Sci. U.S.A. 86:4868-4871(1989).
- 25 [3] Sakane F., Yamada K., Kanoh H., Yokoyama C., Tanabe T. Nature 344:345-348(1990).
 - [4] Ahmed S., Kozma R., Monfries C., Hall C., Lim H.H., Smith P., Lim L. Biochem. J. 272:767-773(1990).
 - [5] Ahmed S., Kozma R., Lee J., Monfries C., Harden N., Lim L. Biochem. J. 280:233-241(1991).
- 30 [6] Ahmed S., Maruyama I.N., Kozma R., Lee J., Brenner S., Lim L. Biochem. J. 287:995-999(1992).
 - [7] Boguski M.S., Bairoch A., Attwood T.K., Michaels G.S. Nature 358:113-113(1992).
 - 932. 3-dehydroquinate synthase (DHQ synthase)

25

5

10



[2] Hawkins AR, Lamb HK; Medline: 96048023 The molecular biology of multidomain proteins. Selected examples." Eur J Biochem 1995;232:7-18.

The 3-dehydroquinate synthase EC:4.6.1.3 domain is present in isolation in various bacterial 3-dehydroquinate synthases and also present as a domain in the pentafunctional AROM polypeptide Swiss:P07547 [2]. 3-dehydroquinate (DHQ) synthase catalyses the formation of dehydroquinate (DHQ) and orthophosphate from 3-deoxy-D-arabino heptulosonic 7 phosphate [1]. This reaction is part of the shikimate pathway which is involved in the biosynthesis of aromatic amino acids.

Number of members: 25

933. Dihydrofolate reductase signature (DiHfolate_red)

Dihydrofolate reductases (EC 1.5.1.3) [1] are ubiquitous enzymes which catalyze the reduction of folic acid into tetrahydrofolic acid. They can be inhibited by a number of antagonists such as trimethroprim and methotrexate which are used as antibacterial or anticancerous agents. A signature pattern was derived from a region in the N-terminal part of these enzymes, which includes a conserved Pro-Trp dipeptide; the tryptophan has been shown [2] to be involved in the binding of substrate by the enzyme.

Consensus pattern[LVAGC]-[LIF]-G-x(4)-[LIVMF]-P-W-x(4,5)-[DE]-x(3)-[FYIV]-x(3)-[STIQ] Sequences known to belong to this class detected by the patternALL, except for type II bacterial, plasmid-encoded, dihydrofolate reductases which do not belong to the same class of enzymes.

- [1] Harpers' Review of Biochemistry, Lange, Los Altos (1985).
- 30 [2] Bolin J.T., Filman D.J., Matthews D.A., Hamlin R.C., Kraut J. J. Biol. Chem. 257:13650-13662(1982).

934. (DIL)

[1] Ponting CP; Medline: 95397417 AF-6/cno: neither a kinesin nor a myosin, but a bit of both." Trends Biochem Sci 1995;20:265-266.

Number of members: 31

5

10

15

20

935. (DNA_gyraseB_C)

DNA topoisomerase II signature (cross-reference = TOPOISOMERASE_II)

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaebacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaebacteria the enzyme, known as DNA gyrase, consists of two subunits (genes gyrA and gyrB [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes parC and parE). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

As a signature pattern for this family of proteins, a region was selected that contains a highly conserved pentapeptide. The pattern is located in gyrB, in parE, and in protein 39 of phage T4 topoisomerase.

20

Consensus pattern [LIVMA]-x-E-G-[DN]-S-A-x-[STAG] Sequences known to belong to this class detected by the pattern ALL.

- 5 [1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).
 - [2] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991).
 - [3] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).
 - [4] Roca J. Trends Biochem. Sci. 20:156-160(1995).
- 10 936. (DNA_topoisolIV)

DNA topoisomerase II signature (cross-reference = TOPOISOMERASE_II)

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaebacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaebacteria the enzyme, known as DNA gyrase, consists of two subunits (genes gyrA and gyrB [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes parC and parE). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

25

30



As a signature pattern for this family of proteins, a region was selected that contains a highly conserved pentapeptide. The pattern is located in gyrB, in parE, and in protein 39 of phage T4 topoisomerase.

751

- 5 Consensus pattern [LIVMA]-x-E-G-[DN]-S-A-x-[STAG] Sequences known to belong to this class detected by the patternALL.
 - [1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).
 - [2] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991).
- 10 [3] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).
 - [4] Roca J. Trends Biochem. Sci. 20:156-160(1995).
 - 937. Prolyl oligopeptidase family serine active site (DPPIV_N_term)
- The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below.
 - Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (Flavobacterium meningosepticum and Aeromonas hydrophila); there is a high degree of sequence conservation between these sequences.
 - Escherichia coli protease II (EC 3.4.21.83) (oligopeptidase B) (gene prtB) which cleaves peptide bonds on the C-terminal side of lysyl and argininyl residues.
 - Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline.
 - Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: STE13) which is responsible for the proteolytic maturation of the alpha-factor precursor.
 - Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: DAP2).
 - Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase). This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus.

5

10

A conserved serine residue has experimentally been shown (in E.coli protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).

Consensus pattern D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for yeast DPAP A.

Note: these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4,E1].

- 15 [1] Rawlings N.D., Polgar L., Barrett A.J. Biochem. J. 279:907-911(1991).
 - [2] Barrett A.J., Rawlings N.D. Biol. Chem. Hoppe-Seyler 373:353-360(1992).
 - [3] Polgar L., Szabo E. Biol. Chem. Hoppe-Seyler 373:361-366(1992).
 - [4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).
- 20 938. Deoxyhypusine synthase (DS)

Eukaryotic initiation factor 5A (eIF-5A) contains an unusual amino acid, hypusine [N epsilon-(4-aminobutyl-2-hydroxy)lysine]. The first step in the post-translational formation of hypusine is catalysed by the enzyme deoxyhypusine synthase (DS) EC:1.1.1.249. The modified version of eIF-5A, and DS, are required for eukaryotic cell proliferation [1]. Number of members:

[1] Liao DI, Wolff EC, Park MH, Davies DR; Medline: 98154315 Crystal structure of the NAD complex of human deoxyhypusine synthase: an enzyme with a ball-and-chain mechanism for blocking the active site." Structure 1998;6:23-32.

25

939. (DUF21)

Many of the sequences in this family are annotated as hemolysins, however this is due to a similarity to Swiss:Q54318 that does not contain this domain. This domain is found in the N-terminus of the proteins adjacent to two intracellular CBS domains CBS.

Number of members: 42

940. (DUF59)

5

This family includes prokaryotic proteins of unknown function. The family also includes PhaH Swiss:O84984 from Pseudomonas putida. PhaH forms a complex with PhaF Swiss:O84982, PhaG Swiss:O84983 and PhaI Swiss:O84985, which hydroxylates phenylacetic acid to 2-hydroxyphenylacetic acid [1]. So members of this family may all be components of ring hydroxylating complexes.

15 Number of members: 15

[1] Olivera ER, Minambres B, Garcia B, Muniz C, Moreno MA, Ferrandez A, Diaz E, Garcia JL, Luengo JM; Medline: 98263372 Molecular characterization of the phenylacetic acid catabolic pathway in Pseudomonas putida U: the phenylacetyl-CoA catabolon." Proc Natl Acad Sci U S A 1998;95:6419-6424.

941. (DUF82)

The protein contains four conserved cysteines that may be involved in metal binding or disulphide bridges.

Number of members: 4

942. Riboflavin kinase / FAD synthetase (FAD_Synth)

This family consists part of the bifunctional enzyme riboflavin kinase / FAD synthetase.

These enzymes have both ATP:riboflavin 5'-phospho transferase and ATP:FMNadenylyltransferase activitys [1]. They catalyse the 5'-phosphorylation of riboflavin to FMN
and the adenylylation of FMN to FAD [1].

25

CAUTION: It is not clear if this region of the enzymes catalyses either or both of the enzymatic reactions.

Number of members: 27

5 [1] Manstein DJ, Pai EF; Medline: 87057286 Purification and characterization of FAD synthetase from Brevibacterium ammoniagenes." J Biol Chem 1986;261:16169-16173.

943. [2Fe-2S] binding domain (fer2_2)

[1] Romao MJ, Archer M, Moura I, Moura JJ, LeGall J, Engh R, Schneider M, Hof P, Huber R; Medline: 96072968 Crystal structure of the xanthine oxidase-related aldehyde oxido-reductase from D. gigas." Science 1995;270:1170-1176.

Number of members: 53

15 944. Filovirus glycoprotein (Filo_glycop)

This family includes an extracellular region from the envelope glycoprotein of Ebola and Marburg viruses. This region is also produced as a separate transcript that gives rise to a non-structural, secreted glycoprotein, which is produced in large amounts and has an unknown function [1]. Processing of this protein may be involved in viral pathogenicity [2].

Number of members: 23

- [1] Volchkov VE, Feldmann H, Volchkova VA, Klenk HD; Medline: 98245155 Processing of the Ebola virus glycoprotein by the proprotein convertase furin." Proc Natl Acad Sci U S A 1998;95:5762-5767.
- [2] Sanchez A, Trappier SG, Mahy BW, Peters CJ, Nichol ST; Medline: 96195018 The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing." Proc Natl Acad Sci U S A 1996;93:3602-3607.
- 30 945. Frataxin-like domain (Frataxin_Cyay)

This family contains proteins that have a domain related to the globular C-terminus of Frataxin the protein that is mutated in Friedreich's ataxia. This domain is found in a family of bacterial proteins. The function of this domain is currently unknown.

20

Number of members:

[1] Gibson TJ, Koonin EV, Musco G, Pastore A, Bork P; Medline: 97084946 Friedreich's ataxia protein: phylogenetic evidence for mitochondrial dysfunction." Trends Neurosci 1996;19:465-468.

946. (GAF)

5

Domain present in phytochromes and cGMP-specific phosphodiesterases.

10 Number of members: 296

[1] Aravind L, Ponting CP; Medline: 98094688 The GAF domain: an evolutionary link between diverse phototransducing proteins." Trends Biochem Sci 1997;22:458-459.

15 947. Galaptin signature (Gal-bind_lectin)

All vertebrates synthesize soluble galactoside-binding lectins [1,2,3] (also known as galectins, galaptins or S-lectin). These carbohydrate-binding proteins are developmentally regulated. Although their exact physiological role is not yet clear they seem to be involved in differentiation, cellular regulation and tissue construction. The sequence of galactoside-binding lectins from electric eel (electrolectin), conger eel (congerin), chicken and a number of mammalian species is known. These lectins are proteins of about 130 to 140 amino acid residues (14 Kd to 16 Kd).

- A number of other proteins are known to belong to this family:
 - Galectin-3 (also known as MAC-2 antigen; CBP-35 or IgE-binding protein), a 35 Kd lectin which binds immunoglobulin E and which is composed of two domains: a N-terminal domain that consist of tandem repeats of a glycine/ proline-rich sequence and a C-terminal galaptin domain.
- Galectin-4 [4], which is composed of two galaptin domains.
 - Galectin-5.
 - Galectin-7 [5], a keratinocyte protein which could be involved in cell-cell and/or cell-matrix interactions necessary for normal growth control.
 - Galectin-8 [6], which is composed of two galaptin domains.

10

15

- Galectin-9 [7], which is composed of two galaptin domains.
- Human eosinophil lysophospholipase (EC 3.1.1.5) [8] (Charcot-Leyden crystal protein), a protein that may have both an enzymatic and a lectin activities. It forms hexagonal bipyramidal crystals in tissues and secretions from sites of eosinophil-associated inflammation.
- Caenorhabditis elegans 32 Kd lactose-binding lectin [9]. This lectin is composed of two galaptin domains.
- Caenorhabditis elegans lec-7 and lec-8.

One of the conserved regions of these lectins contains a tryptophan that has been shown [10] to be essential to the binding of galactosides. This region was used as a signature pattern for these proteins.

Consensus patternW-[GEK]-x-[EQ]-x-[KRE]-x(3,6)-[PCTF]-[LIVMF]-[NQEGSKV]-x-[GH]-x(3)-[DENKHS]-[LIVMFC] [W binds carbohydrate] Sequences known to belong to this class detected by the pattern ALL, except for pig galectin 4.

- [1] Barondes S.H., Gitt M.A., Leffler H., Cooper D.N.W. Biochimie 70:1627-1632(1988).
- [2] Hirabayashi J., Kasai K.-I. J. Biochem. 104:1-4(1988).
- [3] Barondes S.H., Castronovo V., Cooper D.N.W., Cummings R.D., Drickamer K., Feizi
- T., Gitt M.A., Hirabayashi J., Hughes C., Kasai K.-I., Leffler H., Liu F.-T., Lotan R., Mercurio A.M., Monsigny M., Pillair S., Poirer F., Raz A., Rigby P.W.J., Rini J.M., Wang J.L. Cell 76:597-598(1994).
 - [4] Oda Y., Herrmann J., Gitt M., Turck C.W., Burlingame A.L., Barondes S.H., Leffler H. J. Biol. Chem. 268:5929-5939(1993).
- [5] Madsen P., Rasmussen H.H., Flint T., Gromov P., Kruse T.A., Honore B., Vorum H.,Celis J.E. J. Biol. Chem. 270:5823-5829(1995).
 - [6] Hadari Y.R., Paz K., Dekel R., Mestrovic T., Accili D., Zick Y. J. Biol. Chem. 270:3447-3453(1995).
 - [7] Wada J., Kanwar Y.S. J. Biol. Chem. 272:6078-6086(1997).
- [8] Ackerman S.J., Corrette S.E., Rosenberg H.F., Bennett J.C., Mastrianni D.M., Nicholson-Weller A., Weller P.F., Chin D.T., Tenen D.G. J. Immunol. 150:456-468(1993).
 - [9] Hirabayashi J., Satoh M., Kasai K.-I. J. Biol. Chem. 267:15485-15490(1992).
 - [10] Abbott W.M., Feizi T. J. Biol. Chem. 266:5552-5557(1991).

10

948. (GARS) Phosphoribosylglycinamide synthetase signature (phosphoribosylamine glycine ligase)

PROSITE: PDOC00164; cross-reference(s): PS00184

[1] catalyzes the second step in the de novo biosynthesis of purine, the ATP-dependent addition of 5-phosphoribosylamine to glycine to form 5'phosphoribosylglycinamide.

In bacteria GARS is a monofunctional enzyme (encoded by the purD gene), in of a bifunctional enzyme (encoded by the ADE5,7 gene), in higher eukaryotes it is part, with AIRS and with phosphoribosylglycinamide formyltransferase (GART) of a trifunctional enzyme (GARS-AIRS-GART).

The sequence of GARS is well conserved. A highly conserved octapeptide was selected as a signature pattern.

Consensus patternR-F-G-D-P-E-x-[QM]

Sequences known to belong to this class detected by the patternALL.

[1] Aiba A., Mizobuchi K. J. Biol. Chem. 264:21239-21246(1989).

949. GLTT - GLTT repeat (12 copies)

This short repeat of unknown function is found in multiple copies in several C. elegans proteins. The repeat is five residues long and consists of XGLTT where X can be any amino acid. Number of members: 34.

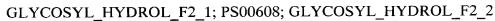
950. Glu synthase - Conserved region in glutamate synthase

- This family represents a region of the glutamate synthase protein. This region is expressed as a seperate subunit in the glutamate synthase alpha subunit from archaebacteria, or part of a large multidomain enzyme in other organisms. The aligned region of these proteins contains a putative FMN binding site and Fe-S cluster. Number of members: 44.
- [1] Medline: 97082505. Sequence of the GLT1 gene from Saccharomyces cerevisiae reveals the domain structure of yeast glutamate synthase. Filetici P, Martegani MP, Valenzuela L, Gonzalez A, Ballario P; Yeast 1996;12:1359-1366.
 - 951. (Glyco_hydro_2) Glycosyl hydrolases family 2 signatures

30

5

10



It has been shown [1,2,E1] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- -Beta-galactosidases (EC 3.2.1.23) from bacteria such as Escherichia coli (genes lacZ and ebgA), Clostridium acetobutylicum, Clostridium thermosulfurogenes, Klebsiella pneumoniae, Lactobacillus delbrueckii, or Streptococcus thermophilus and from the fungi Kluyveromyces lactis.
- -Beta-glucuronidase (EC 3.2.1.31) from Escherichia coli (gene uidA) and from mammals. One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [3], in Escherichia coli lacZ, to be the general acid/base catalyst in the active site of the enzyme. This region has been used as a signature pattern. A highly conserved region located some sixty residues upstream from the active site glutamate has been selected as a second signature pattern.
- Consensus pattern N-x-[LIVMFYWD]-R-[STACN](2)-H-Y-P-x(4)-[LIVMFYWS](2)-x(3)-[DN]-x(2)-G-[LIVMFYW](4) Sequences known to belong to this class detected by the pattern ALL.
 - Consensus pattern [DENQLF]-[KRVW]-N-[HRY]-[STAPPV]-[SAC]-[LIVMFS](3)-W-[GS]-x(2,3)-N-E [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for Rhizobium meliloti lacZ.
 - [1]Henrissat B. Biochem. J. 280:309-316(1991).
 - [2]Schroeder C.J., Robert C., Lenzen G., McKay L.L., Mercenier A. J. Gen. Microbiol.
- 25 137:369-380(1991).
 - [3] Gebler J.C., Aebersold R., Withers S.G. J. Biol. Chem. 267:11126-11130(1992).
 - 952. (Glyco_hydro_3) Glycosyl hydrolases family 3 active site
 - PROSITE: PDOC00621. PROSITE cross-reference(s)PS00775; GLYCOSYL_HYDROL_F3
 - It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:
 - -Beta glucosidases (EC 3.2.1.21) from the fungi Aspergillus wentii (A-3), Hansenula anomala, Kluyveromyces fragilis, Saccharomycopsis fibuligera,(BGL1 and BGL2), Schizophyllum commune and Trichoderma reesei (BGL1).

15

25

30



- -Beta glucosidases from the bacteria Agrobacterium tumefaciens (Cbg1), Butyrivibrio fibrisolvens (bglA), Clostridium thermocellum (bglB), Escherichia coli (bglX), Erwinia chrysanthemi (bgxA) and Ruminococcus albus.
- -Alteromonas strain O-7 beta-hexosaminidase A (EC 3.2.1.52).
- 5 -Bacillus subtilis hypothetical protein yzbA.
 - -Escherichica coli hypothetical protein ycfO and HI0959, the corresponding Haemophilus influenzae protein.

One of the conserved regions in these enzymes is centered on a conserved aspartic acid residue which has been shown [3], in Aspergillus wentii beta-glucosidase A3, to be implicated in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern[LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-x(2)-[SGADNI] [D is the active site residue]

Sequences known to belong to this class detected by the patternALL.

[1]Henrissat B. Biochem. J. 280:309-316(1991).

[2]Castle L.A., Smith K.D., Morris R.O. J. Bacteriol. 174:1478-1486(1992).

[3]Bause E., Legler G. Biochim. Biophys. Acta 626:459-465(1980).

20 953. GP120 - Envelope glycoprotein GP120

The entry of HIV requires interaction of viral GP120 with Swiss:P01730 and a chemokine receptor on the cell surface. Number of members: 17891

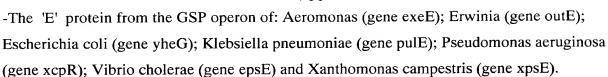
[1]Medline: 98303379. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA; Nature 1998;393:648-659.

954. (GSPII_E) Bacterial type II secretion system protein E signature PROSITE: PDOC00567. PROSITE cross-reference(s) PS00662; T2SP_E

A number of bacterial proteins, some of which are involved in a general secretion pathway (GSP) for the export of proteins (also called the type II pathway) [1,2], have been found to be evolutionary related. These proteins are listed below:

20

5



- -Agrobacterium tumefaciens Ti plasmid virB operon protein 11. This protein is required for the transfer of T-DNA to plants.
- -Bacillus subtilis comG operon protein 1 which is required for the uptake of DNA by competent Bacillus subtilis cells.
- -Aeromonas hydrophila tapB, involved in type IV pilus assembly.
- -Pseudomonas protein pilB, which is essential for the formation of the pili.
- -Pseudomonas aeruginosa protein twitching mobility protein pilT.
 - -Neisseria gonorrhoeae type IV pilus assembly protein pilF.
 - -Vibrio cholerae protein tcpT, which is involved in the biosynthesis of the tcp pilus.
 - -Escherichia coli protein hofB (hopB).
 - -Escherichia coli hypothetical protein ygcB.
 - -Escherichia coli hypothetical protein yggR.

These proteins have from 344 (pilT and virB11) to 568 (tapB) amino acids, they are probably cytoplasmically located and, on the basis of the presence of a conserved P-loop region (see <PDOC00017>), probably bind ATP. A region that overlaps the 'B' motif of ATP-binding proteins was selected as a signature pattern.

Consensus pattern[LIVM]-R-x(2)-P-D-x-[LIVM](3)-G-E-[LIVM]-R-D Sequences known to belong to this class detected by the patternALL, except for ygcB.

[1]Salmond G.P.C., Reeves P.J. Trends Biochem. Sci. 18:7-12(1993).[2]Hobbs M., Mattick J.S. Mol. Microbiol. 10:233-243(1993).

955. (guanylate_cyc) Guanylate cyclases signature

PROSITE: PDOC00425. PROSITE cross-reference(s) PS00452;

GUANYLATE_CYCLASES Guanylate cyclases (EC 4.6.1.2) [1 to 4] catalyze the formation of cyclic GMP (cGMP) from GTP. cGMP acts as an intracellular messenger, activating cGMP dependent kinases and regulating CGMP-sensitive ion channels. The role of cGMP as a second messenger in vascular smooth muscle relaxation and retinal phototransduction is well established. Guanylate cyclase is found both in the soluble and particular

25

30

5

fraction of eukaryotic cells. The soluble and plasma membrane-bound forms differ in structure, regulation and other properties.

Most currently known plasma membrane-bound forms are receptors for small polypeptides. The topology of such proteins is the following: they have a N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain, followed by a large cytoplasmic C-terminal region that can be subdivided into two domains: a protein kinase-like domain that appears important for proper signalling and a cyclase catalytic domain. This topology is schematically represented below.

- 15 The known guanylate cyclase receptors are:
 - -The sea-urchins receptors for speract and resact, which are small peptides that stimulate sperm motility and metabolism.
 - -The receptors for natriuretic peptides (ANF). Two forms of ANF receptors with guanylate cyclase activity are currently known: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which seems to be stimulated more effectively by brain natriuretic peptide (BNP) than by ANP.
 - -The receptor for Escherichia coli heat-stable enterotoxin (GC-C). The endogenous ligand for this intestinal receptor seems to be a small peptide called guanylin.
 - -Retinal guanylate cyclase (retGC) which probably plays a specific functional role in the rods and/or cones of photoreceptors. It is not known if this protein acts as receptor, but its structure is similar to that of the other plasma membrane-bound GCs.

The soluble forms of guanylate cyclase are cytoplasmic heterodimers. The two subunits, alpha and beta are proteins of from 70 to 82 Kd which are highly related. Two forms of beta subunits are currently known: beta-1 which seems to be expressed in lung and brain, and beta-2 which is more abundant in kidney and liver.

The membrane and cytoplasmic forms of guanylate cyclase share a conserved domain which is probably important for the catalytic activity of the enzyme. Such a domain is also found twice in the different forms of membrane-bound adenylate cyclases (also known as

25

5

class-III) [5,6] from mammals, slime mold or Drosophila. A consensus pattern was derived from the most conserved region in that domain.

Consensus patternG-V-[LIVM]-x(0,1)-G-x(5)-[FY]-x-[LIVM]-[FYW]-[GS]-[DNTHKW]-[DNT]-[IV]-[DNTA]-x(5)-[DE]

Sequences known to belong to this class detected by the patternALL, except for the sea urchin Arbacia punctulata resact receptor which lack this domain.

Note this pattern will detect both domains of adenylate cyclases class-III.

- 10 [1]Koesling D., Boehme E., Schultz G. FASEB J. 5:2785-2791(1991).
 - [2]Garbers D.L. New Biol. 2:499-504(1990).
 - [3]Garbers D.L. Cell 71:1-4(1992).
 - [4] Yuen P.S.T., Garbers D.L. Annu. Rev. Neurosci. 15:193-225(1992).
 - [5] Iyengar R. FASEB J. 7:768-775(1993).
- 15 [6]Barzu O., Danchin A. Prog. Nucleic Acid Res. Mol. Biol. 49:241-283(1994).
 - 956. Hemolysin-type calcium-binding region signature (HemolysinCabinD)
 - Gram-negative bacteria produce a number of proteins which are secreted into the growth medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, seem [1] to share two properties: they bind calcium and they contain a variable number of tandem repeats consisting of a nine amino acid motif rich in glycine, aspartic acid and asparagine. It has been shown [2] that such a domain is involved in the binding of calcium ions in a parallel beta roll structure. The proteins which are currently known to belong to this category are:
 - Hemolysins from various species of bacteria. Bacterial hemolysins are exotoxins that attack blood cell membranes and cause cell rupture. The hemolysins which are known to contain such a domain are those from: E. coli (gene hlyA), A. pleuropneumoniae (gene appA), A. actinomycetemcomitans and P. haemolytica (leukotoxin) (gene lktA).
- Cyclolysin from Bordetella pertussis (gene cyaA). A multifunctional protein which is both an adenylate cyclase and a hemolysin.
 - Extracellular zinc proteases: serralysin (EC 3.4.24.40) from Serratia, prtB and prtC from Erwinia chrysanthemi and aprA from Pseudomonas aeruginosa.
 - Nodulation protein nodO from Rhizobium leguminosarum.

25

A signature pattern was derived from conserved positions in the sequence of the calciumbinding domain.

Consensus pattern D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D Sequences known to belong to this class detected by the pattern ALL.

Note: This pattern is found once in nodO and the extracellular proteases but up to 5 times in some hemolysin/cyclolysins.

- [1] Economou A., Hamilton W.D.O., Johnston A.W.B., Downie J.A. EMBO J. 9:349-354(1990).
 - [2] Baumann U., Wu S., Flaherty K.M., McKay D.B. EMBO J. 12:3357-3364(1993).
 - 957. Hint module (Hint)

This is an alignment of the Hint module in the Hedgehog proteins. It does not include any Inteins which also possess the Hint module.

Number of members: 36

- [1] Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ; Medline: 97474313 Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins." Cell 1997;91:85-97.
 - 958. Hydantoinase/oxoprolinase (Hydantoinase)
 - This family includes the enzymes hydantoinase and oxoprolinase EC:3.5.2.9. Both reactions involve the hydrolysis of 5-membered rings via hydrolysis of their internal imide bonds [1]. Number of members: 14
- [1] Ye GJ, Breslow EB, Meister A, Guo-jie GE\$[corrected to Ye GJ]; Medline: 97113037 The amino acid sequence of rat kidney 5-oxo-L-prolinase determined by cDNA cloning" [published erratum appears in J Biol Chem 1997 Feb 14;272(7):4646] J Biol Chem 1996;271:32293-32300.

20

5

C 1

959. IMP dehydrogenase / GMP reductase signature (IMPDH_N)

IMP dehydrogenase (EC 1.1.1.205) (IMPDH) catalyzes the rate-limiting reaction of de novo GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [1]. Inhibition of IMP dehydrogenase activity results in the cessation of DNA synthesis. As IMP dehydrogenase is associated with cell proliferation, it is a possible target for cancer chemotherapy. Mammalian and bacterial IMPDHs are tetramers of identical chains. There are two IMP dehydrogenase isozymes in humans [2].

GMP reductase (EC 1.6.6.8) catalyzes the irreversible and NADPH-dependent reductive deamination of GMP into IMP [3]. It converts nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and maintains intracellular balance of A and G nucleotides.

IMP dehydrogenase and GMP reductase share many regions of sequence similarity. One of these regions is centered on a cysteine residue thought [3] to be involved in binding IMP. This region was used as a signature pattern.

Consensus pattern[LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T [C is the putative IMP-binding residue] Sequences known to belong to this class detected by the pattern ALL.

- [1] Collart F.R., Huberman E. J. Biol. Chem. 263:15769-15772(1988).
- [2] Natsumeda Y., Ohno S., Kawasaki H., Konno Y., Weber G., Suzuki K. J. Biol. Chem. 265:5292-5295(1990).
- 25 [3] Andrews S.C., Guest J.R. Biochem. J. 255:35-43(1988).

960. impB/mucB/samB family (IMS)

These proteins are involved in UV protection (Swiss).

30 Number of members: 38

961. Type II intron maturase (Intron maturas2)

25

30

5

Group II introns use intron-encoded reverse transcriptase, maturase and DNA endonuclease activities for site-specific insertion into DNA [2]. Although this type of intron is self splicing in vitro they require a maturase protein for

splicing in vivo. It has been shown that a specific region of the aI2 intron is needed for the maturase function [1]. This region was found to be conserved in group II introns and called domain X [3].

Number of members: 335

- [1] Moran JV, Mecklenburg KL, Sass P, Belcher SM, Mahnke D, Lewin A, Perlman P;
 Medline: 94301788 Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron al2. Nucleic Acids Res 1994;22:2057-2064.
 - [2] Guo H, Zimmerly S, Perlman PS, Lambowitz AM; Medline: 98031910 Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA." EMBO J 1997;16:6835-6848.
 - [3] Mohr G, Perlman PS, Lambowitz AM; Medline: 94077696 Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function." Nucleic Acids Res 1993;21:4991-4997.
- 20 962. LAGLIDADG endonuclease (Intron_maturase)
 - [1] Heath PJ, Stephens KM, Monnat RJ Jr, Stoddard BL; Medline: 97331323 The structure of I-Crel, a group I intron-encoded homing endonuclease." Nat Struct Biol 1997;4:468-476.
 - [2] Belfort M, Roberts RJ; Medline: 97402526 Homing endonucleases: keeping the house in order." Nucleic Acids Res 1997;25:3379-3388.
 - [3] Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS; Medline: 98026854 Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family." Nucleic Acids Res 1997;25:4626-4638.

Number of members:

220

963. Isopentenyl transferase (IPT)

30

5



Isopentenyl transferase / dimethylallyl transferase synthesizes isopentenyladensosine 5'-monophosphate, a cytokinin that induces shoot formation on host plants infected with the Ti plasmid [1].

Number of members: 1

16

- [1] Canaday J, Gerad JC, Crouzet P, Otten L; Medline: 93101133 "Organization and functional analysis of three T-DNAs from the vitopine Ti plasmid pTiS4." Mol Gen Genet 1992;235:292-303.
- 10 964. Laminin EGF-like (Domains III and V) (laminin_EGF)

This family is like EGF but has 8 conserved cysteines instead of 6.

Number of members: 501

- [1] Engel J; Medline: 93041759 Laminins and other strange proteins." Biochemistry 1992;31:10643-10651.
 - 965. Legume lectins signatures (lectin_legA)
- Leguminous plants synthesize sugar-binding proteins which are called legume lectins [1,2]. These lectins are generally found in the seeds. The exact function of legume lectins is not known but they may be involved in the attachment of nitrogen-fixing bacteria to legumes and in the protection against pathogens. Legume lectins bind calcium and manganese (or other transition metals).

Legume lectins are synthesized as precursor proteins of about 230 to 260 amino acid residues. Some legume lectins are proteolytically processed to produce two chains: beta (which corresponds to the N-terminal) and alpha (C-terminal). The lectin concanavalin A (conA) from jack bean is exceptional in that the two chains are transposed and ligated (by formation of a new peptide bond). The N-terminus of mature conA thus corresponds to that of the alpha chain and the C-terminus to the beta chain.

Two signature patterns were developed specific to legume lectins: the first is located in the C-terminal section of the beta chain and contains a conserved aspartic acid residue important for

the binding of calcium and manganese; the second one is located in the N-terminal of the

20

25

30

5



alpha chain.

Consensus pattern [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST] [D binds manganese and

calcium] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIVF]-G-[LF]-[ST] Sequences known to belong to this class detected by the pattern ALL.

- 10 [1] Sharon N., Lis H. FASEB J. 4:3198-320(1990).
 - [2] Lis H., Sharon N. Annu. Rev. Biochem. 55:33-37(1986).
 - 966. Malate synthase signature (malate_synthase)
- Malate synthase (EC 4.1.3.2) catalyzes the aldol condensation of glyoxylate with acetyl-CoA to form malate the second step of the glyoxylate bypass, an alternative to the tricarboxylic acid cycle in bacteria, fungi and plants. Malate synthase is a protein of 530 to 570 amino acids whose sequence is highly conserved across species [1]. As a signature pattern, a very conserved region was selected in the central section of the enzyme.

Consensus pattern[KR]-[DENQ]-H-x(2)-G-L-N-x-G-x-W-D-Y-[LIVM]-F Sequences known to belong to this class detected by the pattern ALL.

- [1] Bruinenberg P.G., Blaauw M., Kazemier B., Ab G. Yeast 6:245-254(1990).
- 967. MatK/TrnK amino terminal region (MatK_N)
 - [1] Mohr G, Perlman PS, Lambowitz AM; Medline: 94077696 Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function." Nucleic Acids Res 1993;21:4991-4997.

Number of members: 495

968. MOZ/SAS family (MOZ_SAS)

This region of these proteins has been suggested to be homologous to acetyltransferases [1]. However the similarity is not supported by standard sequence analysis.

Number of members: 15

5

- [1] Kamine J, Elangovan B, Subramanian T, Coleman D, Chinnadurai G; Medline: 96182937 Identification of a cellular protein that specifically interacts with the essential cysteine region of the HIV-1 Tat transactivator." Virology 1996;216:357-366.
- [2] Reifsnyder C, Lowell J, Clarke A, Pillus L; Medline: 96376969 Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases" [see comments] [published erratum appears in Nat Genet 1997 May;16(1):109] Nat Genet 1996;14:42-49.

969. mRNA capping enzyme (mRNA_cap_enzyme)

15

10

[1] Hakansson K, Doherty AJ, Shuman S, Wigley DB; Medline: 97304383 X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes." Cell 1997;89:545-553.

20

30

7

Number of members:

970. DNA mismatch repair proteins mutS family signature (MutS C)

Mismatch repair contributes to the overall fidelity of DNA replication [1]. It involves the correction of mismatched base pairs that have been missed by the proofreading element of the DNA polymerase complex. The sequence of some proteins involved in mismatch repair in different organisms have been found to be evolutionary related [2,3]. One of these families is called mutS [4,E1], it consists of:

- Prokaroytic protein mutS protein (also called hexA in Streptococcus pneumoniae). Muts is thought to carry out the mismatch recognition step of DNA repair.
- Eukaryotic MSH1, which is involved in mitochondrial DNA repair.
- Eukaryotic MSH2, which is involved in nuclear postreplication mismatch repair. MSH2 heterodimerizes with MSH6. In man, MSH2 is involved in a form of familial hereditary nonpolyposis colon cancer (HNPCC).

20

25

30

5



- Eukaryotic MSH3, which is probably involved in the repair of large loops.
- Eukaryotic MSH4, which is involved in meiotic recombination.
- Eukaryotic MSH5, which is involved in meiotic recombination.
- Eukaryotic MSH6 (also known as G/T mismatch binding protein), a DNA-repair protein that binds to G/T mismatches through heterodimerization with MSH2.
- Prokaryotic protein mutS2 whose function is not yet known.
- A coral (Sarcophyton glaucum) mitochondrial encoded mutS-like protein.

As a signature pattern for this class of mismatch repair proteins a region rich in glycine and negatively charged residues was selected This region is found

in the C-terminal section of these proteins; about 80 residues to the C-terminal of an ATP-binding site motif 'A' (P-loop) (see <PDOC00017>).

Consensus pattern[ST]-[LIVMF]-x-[LIVM]-x-D-E-[LIVMFY]-[GC]-[RKH]-G-[GST]- x(4)-G Sequences known to belong to this class detected by the pattern ALL, except for mutS2.

- [1] Modrich P. Annu. Rev. Biochem. 56:435-466(1987).
- [2] Haber L.T., Walker G.C. EMBO J. 10:2707-2715(1991).
- [3] New L., Liu K., Crouse G.F. Mol. Gen. Genet. 239:97-108(1993).
- [4] Eisen J.A. Nucleic Acids Res. 26:4291-4300(1998).

971. MutS family, N-terminal putative DNA binding domain (MutS_N)

This family consists of the N-terminal region of proteins in the mutS family of DNA mismatch repair proteins and is found associated with MutS_C located in the C-terminal region. The mutS family of proteins is named after the salmonella typhimurium MutS protein involved in mismatch repair; other members of the family included the eukaryotic MSH 1,2,3,4,5 and 6 proteins. These have various roles in DNA repair and recombination. Human MSH has been implicated in non-polyposis colorectal carcinoma (HNPCC) and is a mismatch binding protein [2]. The aligned region corresponds in part with domains A1, A2 (which may bind DNA) and B (which binds dsDNA in vitro) from T. thermophilus MutS as characterised in [1].

Number of members: 43

972. Domain in Myosin and Kinesin Tails (MyTH4)

15

20

25

Domain present twice in myosin-VIIa, and also present in 3 other myosins.

[1] Chen ZY, Hasson T, Kelley PM, Schwender BJ, Schwartz MF, Ramakrishnan M,
 Kimberling WJ, Mooseker MS, Corey DP; Medline: 97038686 Molecular cloning and domain structure of human myosin-VIIa, the gene product defective in Usher syndrome 1B."
 Genomics 1996;36:440-448.

Number of members:

21

973. Sodium and potassium ATPases beta subunits signatures (Na K-ATPase)

The sodium pump (Na+,K+ ATPase), located in the plasma membrane of all animal cells [1], is an heterotrimer of a catalytic subunit (alpha chain), a glycoprotein subunit of about 34 Kd (beta chain) and a small hydrophobic protein of about 6 Kd. The beta subunit seems [2] to regulate, through the assembly of alpha/beta heterodimers, the number of sodium pumps transported to the plasma membrane.

Structurally the beta subunit is composed of a charged cytoplasmic domain of about 35 residues, followed by a transmembrane region, and a large extracellular domain that contains three disulfide bonds and glycosylation sites. This structure is schematically represented in the figure below.

'C': conserved cysteine involved in a disulfide bond.

'*': position of the patterns.

Two isoforms of the beta subunit (beta-1 and beta-2) are currently known; they share about 50% sequence identity. Gastric (K+, H+) ATPase (proton pump) responsible for acid production in the stomach consist of two subunits [3]; the beta chain is highly similar to the sodium pump beta subunits. Two signature patterns were developed for beta subunits. The

30

5

10

first is located in the cytoplasmic domain, while the second is found in the extracellular domain and contains two of the cysteines involved in disulfide bonds.

Consensus pattern [FYW]-x(2)-[FYW]-x-[FYW]-[DN]-x(6)-[LIVM]-G-R-T-x(3)-W Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [RK]-x(2)-C-[RKQWI]-x(5)-L-x(2)-C-[SA]-G [The two C's are involved in disulfide bonds] Sequences known to belong to this class detected by the patternALL, except for the beta subunit of the sodium pump of brine shrimp whose sequence is highly divergent in that region.

- [1] Horisberger J.D., Lemas V., Krahenbul J.P., Rossier B.C. Annu. Rev. Physiol. 53:565-584(1991).
- [2] McDonough A.A., Gerring K., Farley R.A. FASEB J. 4:1598-1605(1990).
- [3] Toh B.-H., Gleeson P.A., Simpson R.J., Moritz R.L., Callaghan J.M., Goldkorn I., Jones C.M., Martinelli T.M., Mu F.-T., Humphris D.C., Pettitt J.M., Mori Y., Masuda T., Sobieszczuk P., Weinstock J., Mantamadiotis T., Baldwin G.S. Proc. Natl. Acad. Sci. U.S.A. 87:6418-6422(1990).
- 974. Respiratory-chain NADH dehydrogenase subunit 1 signatures (NADHdh)

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there are fifteen which are located in the membrane part, seven of which are encoded by the mitochondrial and chloroplast genomes of most species. The most conserved of these organelle-encoded subunits is known as subunit 1 (gene ND1 in mitochondrion, and NDH1 in chloroplast) and seems to contain the ubiquinone binding site.

The ND1 subunit is highly similar to subunit 4 of Escherichia coli formate hydrogenlyase (gene hycD), subunit C of hydrogenase-4 (gene hyfC). Paracoccus denitrificans NQO8 and Escherichia coli nuoH NADH-ubiquinone oxidoreductase subunits also belong to this family [3]. Two signature patterns were developed based on conserved regions of this subunit.

25

30

5

Consensus pattern G-[LIVMFYKRS]-[LIVMAGP]-Q-x-[LIVMFY]-x-D-[AGIM]- [LIVMFTA]- K-[LVMYST]-[LIVMFYG]-x-[KR]-[EQG] Sequences known to belong to this class detected by the patternALL, except for watermelon and Leishmania ND1.

Consensus pattern P-F-D-[LIVMFYQ]-[STAGPVM]-E-[GAC]-E-x-[EQ]-[LIVMS]-x(2)-G Sequences known to belong to this class detected by the pattern ALL, except for Chlamydomonas reinhardtii and Pisaster ochraceus ND1, and tobacco NDH1.

- 10 [1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).
 - [2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).
 - [3] Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).
- 975. Nickel-dependent hydrogenases large subunit signatures (NiFeSe_Hases)

Hydrogenases are enzymes that catalyze the reversible activation of hydrogen and which occur widely in prokaryotes as well as in some eukaryotes. There are various types of hydrogenases, but all of them seem to contain at least one iron-sulfur cluster. They can be broadly divided into two groups: hydrogenases containing nickel and, in some cases, also selenium (the [NiFe] and [NiFeSe] hydrogenases) and those lacking nickel (the [Fe] hydrogenases).

The [NiFe] and [NiFeSe] hydrogenases are heterodimer that consist of a small subunit that contains a signal peptide and a large subunit. All the known large subunits seem to be evolutionary related [1]; they contain two Cys-x-x- Cys motifs; one at their N-terminal end; the other at their C-terminal end. These four cysteines are involved in the binding of nickel [2]. In the [NiFeSe] hydrogenases the first cysteine of the C-terminal motif is a selenocysteine which has experimentally been shown to be a nickel ligand [3]. Two patterns were developed which are centered on the Cys-x-x-Cys motifs.

Alcaligenes eutrophus possess a NAD-reducing cytoplasmic hydrogenase (hoxS) [4]; this enzyme is composed of four subunits. Two of these subunits (beta and delta) are responsible for the hydrogenase reaction and are evolutionary related to the large and small subunits of

20

membrane-bound hydrogenases. The alpha subunit of coenzyme F420 hydrogenase (EC 1.12.99.1) (FRH) from archaebacterial methanogens also belongs to this family.

Consensus pattern R-G-[LIVMF]-E-x(15)-[QESM]-R-x-C-G-[LIVM]-C [The two C's are nickel ligands] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [FY]-D-P-C-[LIM]-[ASG]-C-x(2,3)-H [The two C's are nickel ligands] Sequences known to belong to this class detected by the pattern ALL.

- [1] Menon N.K., Robbins J., Peck H.D. Jr., Chatelus C.Y., Choi E.-S., Przybyla A.E. J. Bacteriol. 172:1969-1977(1990).
 - [2] Volbeda A., Charon M.-H., Piras C., Hatchikian E.C., Frey M., Fontecilla-Camps J.C. Nature 373:580-587(1995).
 - [3] Eidsness M.K., Scott R.A., Prickrill B., der Vartaninan D.V., LeGall J., Moura I., Moura J.J.G., Peck H.D. Jr. Proc. Natl. Acad. Sci. U.S.A. 86:147-151(1989).
 - [4] Tran-Betcke A., Warnecke U., Boecker C., Zaborosch C., Friedrich B. J. Bacteriol. 172:2920-2929(1990).
 - 976. NADH-Ubiquinone oxidoreductase (complex I), chain 5 C-terminus (oxidored_q1_C)
 - This sub-family represents a carboxyl terminal extension of oxidored_q1. Only NADH-Ubiquinone chain 5 from chloroplasts are in this family. This sub-family is part of complex I which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane.
- Number of members: 572
 - [1] Walker JE; Medline: 93110040 The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains." Q Rev Biophys 1992;25:253-324.
- 977. NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus (oxidored_q1_N)
 - This sub-family represents an amino terminal extension of oxidored_q1. Only NADH-Ubiquinone chain 5 and eubacterial chain L are in this family. This sub-family is part of

15

20

25



complex I which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane.

774

Number of members: 546

5 [1] Walker JE; Medline: 93110040 The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains." Q Rev Biophys 1992;25:253-324.

978. oxidored_q2. NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4L (EC 1.6.5.3). ND4L OR NAD4L. Arabidopsis thaliana (Mouse-ear cress). Mitochondrion. OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.

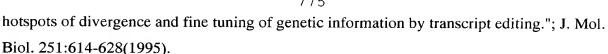
- [1] SEQUENCE FROM N.A. MEDLINE; 93156682. Brandt P., Sunkel S., Unseld M., Brennicke A., Knoop V.; "The nad4L gene is encoded between exon c of nad5 and orf25 in the Arabidopsis mitochondrial genome."; Mol. Gen. Genet. 236:33-38(1992).
 [2] SEQUENCE FROM N.A. STRAIN=CV. COLUMBIA; MEDLINE; 97141919 Unseld M., Marienfeld J.R., Brandt P., Brennicke A.; "The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides."; Nat. Genet. 15:57-61(1997).
- 979. oxidored_q4. Protein name NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 3, CHLOROPLAST. Synonym(s)EC 1.6.5.3. Gene name(s)NDHC OR NDH3 From Zea mays (Maize) Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.

SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.

- [1] SEQUENCE FROM N.A. MEDLINE; 89281491. Steinmueller K., Ley A.C., Steinmetz
 A.A., Sayre R.T., Bogorad L.; "Characterization of the ndhC-psbG-ORF157/159 operon of maize plastid DNA and of the cyanobacterium Synechocystis sp. PCC6803."; Mol. Gen. Genet. 216:60-69(1989).
 - [2] SEQUENCE FROM N.A. MEDLINE; 95395841. Maier R.M., Neckermann K., Igloi G.L., Koessel H.; "Complete sequence of the maize chloroplast genome: gene content,

20

25



980. PAC: PAC motif

- 5 PAC motif occurs C-terminal to a subset of all known PAS motifs. It is proposed to contribute to the PAS domain fold [3]. Number of members: 181
 - [1] Medline: 97446881 PAS domain S-boxes in archaea, bacteria and sensors for oxygen and redox. Zhulin IB, Taylor BL, Dixon R; Trends Biochem Sci 1997;22:331-333.
- 10 [2] Medline: 95275818. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. Borgstahl GE, Williams DR, Getzoff ED; Biochemistry 1995;34:6278-6287.
 - [3] Medline: 98044337. PAS: a multifunctional domain family comes to light. Ponting CP, Aravind L; Curr Biol 1997;7:674-677.

981. PARP: Poly(ADP-ribose) polymerase catalytic region.

Poly(ADP-ribose) polymerase catalyses the covalent attachment of ADP-ribose units from NAD+ to itself and to a limited number of other DNA binding proteins, which decreases their affinity for DNA. Poly(ADP-ribose) polymerase is a regulatory component induced by DNA damage.

The carboxyl-terminal region is the most highly conserved region of the protein. Experiments have shown that a carboxyl 40 kDa fragment is still catalytically active [2]. Number of members: 19

- [1] Medline: 96353841 Structure of the catalytic fragment of poly(AD-ribose) polymerase from chicken. Ruf A, Mennissier de Murcia J, de Murcia G, Schulz GE; Proc Natl Acad Sci USA 1996;93:7481-7485.
- [2] Medline: 93293867 The carboxyl-terminal domain of human poly(ADP-ribose) 30 polymerase. Overproduction in Escherichia coli, large scale purification, and characterization. Simonin F, Hofferer L, Panzeter PL, Muller S, de Murcia G, Althaus FR; J Biol Chem 1993;268:13454-13461.
 - 982. PC rep: Proteasome/cyclosome repeat

[1] Medline: 97348748 A repetitive sequence in subunits of the 26S proteasome and 20S cyclosome (anaphase-promoting complex). Lupas A, Baumeister W, Hofmann K; Trends Biochem Sci 1997;22:195-196.

Number of members:

112

5

983. Peptidase_M1: Peptidase family M1

Members of this family are aminopeptidases. The members differ widely in specificity, hydrolysing acidic, basic or neutral N-terminal residues. This family includes leukotriene-A4 hydrolase Swiss:P09960, this enzyme also has an aminopeptidase activity [1]. Number of

10 members: 72

[1] Medline: 95405261 Evolutionary families of metallopeptidases. Rawlings ND, Barrett AJ; Meth Enzymol 1995;248:183-228.

984. Neutral zinc metallopeptidases, zinc-binding region signature (Peptidase_M8)
PROSITE cross-reference(s) PS00142; ZINC_PROTEASE

The majority of zinc-dependent metallopeptidases (with the notable exception of the carboxypeptidases) share a common pattern of primary structure [1,2,3] in the part of their sequence involved in the binding of zinc, and can be grouped together as a superfamily, known as the metzincins, on the basis of this sequence similarity. They can be classified into a number of distinct families [4,E1] which are listed below along with the proteases which are currently known to belong to these families.

Family M1

- Bacterial aminopeptidase N (EC 3.4.11.2) (gene pepN).
 - Mammalian aminopeptidase N (EC 3.4.11.2).
 - Mammalian glutamyl aminopeptidase (EC 3.4.11.7) (aminopeptidase A). It may play a role in regulating growth and differentiation of early B-lineage cells.
 - Yeast aminopeptidase yscII (gene APE2).
- 30 Yeast alanine/arginine aminopeptidase (gene AAP1).
 - Yeast hypothetical protein YIL137c.
 - Leukotriene A-4 hydrolase (EC 3.3.2.6). This enzyme is responsible for the hydrolysis of an epoxide moiety of LTA-4 to form LTB-4; it has been shown that it binds zinc and is capable of peptidase activity.

20

30

Family M2

- Angiotensin-converting enzyme (EC 3.4.15.1) (dipeptidyl carboxypeptidase I) (ACE) the enzyme responsible for hydrolyzing angiotensin I to angiotensin II. There are two forms of ACE: a testis-specific isozyme and a somatic isozyme which has two active centers.
- 5 Family M3
 - Thimet oligopeptidase (EC 3.4.24.15), a mammalian enzyme involved in the cytoplasmic degradation of small peptides.
 - Neurolysin (EC 3.4.24.16) (also known as mitochondrial oligopeptidase M or microsomal endopeptidase).
- Mitochondrial intermediate peptidase precursor (EC 3.4.24.59) (MIP). It is involved the second stage of processing of some proteins imported in the mitochondrion.
 - Yeast saccharolysin (EC 3.4.24.37) (proteinase yscD).
 - Escherichia coli and related bacteria dipeptidyl carboxypeptidase (EC 3.4.15.5) (gene dcp).
 - Escherichia coli and related bacteria oligopeptidase A (EC 3.4.24.70) (gene opdA or prlC).
 - Yeast hypothetical protein YKL134c.

Family M4

- Thermostable thermolysins (EC 3.4.24.27), and related thermolabile neutral proteases (bacillolysins) (EC 3.4.24.28) from various species of Bacillus.
- Pseudolysin (EC 3.4.24.26) from Pseudomonas aeruginosa (gene lasB).
- Extracellular elastase from Staphylococcus epidermidis.
- Extracellular protease prt1 from Erwinia carotovora.
- Extracellular minor protease smp from Serratia marcescens.
- Vibriolysin (EC 3.4.24.25) from various species of Vibrio.
- Protease prtA from Listeria monocytogenes.
 - Extracellular proteinase proA from Legionella pneumophila.

Family M5

- Mycolysin (EC 3.4.24.31) from Streptomyces cacaoi.

Family M6

- Immune inhibitor A from Bacillus thuringiensis (gene ina). Ina degrades two classes of insect antibacterial proteins, attacins and cecropins.

15

20

Family M7

- Streptomyces extracellular small neutral proteases

Family M8

- Leishmanolysin (EC 3.4.24.36) (surface glycoprotein gp63), a cell surface protease from various species of Leishmania.

Family M9

- Microbial collagenase (EC 3.4.24.3) from Clostridium perfringens and Vibrio alginolyticus.

Family M10A

- Serralysin (EC 3.4.24.40), an extracellular metalloprotease from Serratia.
- Alkaline metalloproteinase from Pseudomonas aeruginosa (gene aprA).
- Secreted proteases A, B, C and G from Erwinia chrysanthemi.
- Yeast hypothetical protein YIL108w.

Family M10B

- Mammalian extracellular matrix metalloproteinases (known as matrixins) [5]: MMP-1 (EC 3.4.24.7) (interstitial collagenase), MMP-2 (EC 3.4.24.24) (72 Kd gelatinase), MMP-9 (EC 3.4.24.35) (92 Kd gelatinase), MMP-7 (EC 3.4.24.23) (matrylisin), MMP-8 (EC 3.4.24.34) (neutrophil collagenase), MMP-3 (EC 3.4.24.17) (stromelysin-1), MMP-10 (EC 3.4.24.22) (stromelysin-2), and MMP-11 (stromelysin-3), MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).
- Sea urchin hatching enzyme (envelysin) (EC 3.4.24.12). A proteas that allows the embryo to digest the protective envelope derived from the egg extracellular matrix.
 - Soybean metalloendoproteinase 1.

Family M11

- Chlamydomonas reinhardtii gamete lytic enzyme (GLE).

Family M12A

- Astacin (EC 3.4.24.21), a crayfish endoprotease.
- Meprin A (EC 3.4.24.18), a mammalian kidney and intestinal brush border

15

20

metalloendopeptidase.

- Bone morphogenic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity. The Drosophila homolog of BMP-1 is the dorsal-ventral patterning protein tolloid.
- Blastula protease 10 (BP10) from Paracentrotus lividus and the related protein SpAN from Strongylocentrotus purpuratus.
 - Caenorhabditis elegans protein toh-2.
 - Caenorhabditis elegans hypothetical protein F42A10.8.
 - Choriolysins L and H (EC 3.4.24.67) (also known as embryonic hatching proteins LCE and HCE) from the fish Oryzias lapides. These proteases participates in the breakdown of the egg envelope, which is derived from the egg extracellular matrix, at the time of hatching.

Family M12B

- Snake venom metalloproteinases [6]. This subfamily mostly groups proteases that act in hemorrhage. Examples are: adamalysin II (EC 3.4.24.46), atrolysin C/D (EC 3.4.24.42), atrolysin E (EC 3.4.24.44), fibrolase (EC 3.4.24.72), trimerelysin I (EC 3.4.25.52) and II (EC 3.4.25.53).
- Mouse cell surface antigen MS2.

Family M13

- Mammalian neprilysin (EC 3.4.24.11) (neutral endopeptidase) (NEP).
- Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), which process the precursor of endothelin to release the active peptide.
- Kell blood group glycoprotein, a major antigenic protein of erythrocytes. The Kell protein is very probably a zinc endopeptidase.
 - Peptidase O from Lactococcus lactis (gene pepO).

Family M27

- Clostridial neurotoxins, including tetanus toxin (TeTx) and the various botulinum toxins (BoNT). These toxins are zinc proteases that block neurotransmitter release by proteolytic cleavage of synaptic proteins such as synaptobrevins, syntaxin and SNAP-25 [7,8].

15

20

25

30

Family M30

- Staphylococcus hyicus neutral metalloprotease.

Family M32

- Thermostable carboxypeptidase 1 (EC 3.4.17.19) (carboxypeptidase Taq), an enzyme from Thermus aquaticus which is most active at high temperature.

Family M34

- Lethal factor (LF) from Bacillus anthracis, one of the three proteins composing the anthrax toxin.

Family M35

- Deuterolysin (EC 3.4.24.39) from Penicillium citrinum and related proteases from various species of Aspergillus.

Family M36

- Extracellular elastinolytic metalloproteinases from Aspergillus.

From the tertiary structure of thermolysin, the position of the residues acting as zinc ligands and those involved in the catalytic activity are known. Two of the zinc ligands are histidines which are very close together in the sequence; C-terminal to the first histidine is a glutamic acid residue which acts as a nucleophile and promotes the attack of a water molecule on the carbonyl carbon of the substrate. A signature pattern which includes the two histidine and the glutamic acid residues is sufficient to detect this superfamily of proteins.

Consensus pattern[GSTALIVN]-x(2)-H-E-[LIVMFYW]-{DEHRKP}-H-x-[LIVMFYWGSPQ]

[The two H's are zinc ligands] [E is the active site residue]

Sequences known to belong to this class detected by the patternALL, except for members of families M5, M7 amd M11.

Other sequence(s) detected in SWISS-PROT57; including Neurospora crassa conidiation-specific protein 13 which could be a zinc-protease.

[1] Jongeneel C.V., Bouvier J., Bairoch A. FEBS Lett. 242:211-214(1989).

25

30



[2] Murphy G.J.P., Murphy G., Reynolds J.J. FEBS Lett. 289:4-7(1991).

[3]Bode W., Grams F., Reinemer P., Gomis-Rueth F.-X., Baumann U., McKay D.B., Stoecker W. Zoology 99:237-246(1996).

- [4] Rawlings N.D., Barrett A.J. Meth. Enzymol: 248:183-228(1995).
- [5] Woessner J. Jr. FASEB J. 5:2145-2154(1991). 5
 - [6]Hite L.A., Fox J.W., Bjarnason J.B. Biol. Chem. Hoppe-Seyler 373:381-385(1992).
 - [7] Montecucco C., Schiavo G. Trends Biochem. Sci. 18:324-327(1993).
 - [8] Niemann H., Blasi J., Jahn R. Trends Cell Biol. 4:179-185(1994).
- 985. PHO4: Phosphate transporter family 10 This family includes PHO-4 from Neurospora crassa which is a is a Na(+)-phosphate symporter [1]. This family also contains the leukemia virus receptor Swiss:Q08344. Number of members: 41
- [1] Medline: 95249577 Repressible cation-phosphate symporters in Neurospora crassa. 15 Versaw WK, Metzenberg RL; Proc Natl Acad Sci U S A 1995;92:3884-3887.
 - 986. Photosynthetic reaction center proteins signature (photoRC) PROSITE cross-reference(s): PS00244; REACTION_CENTER

In the photosynthetic reaction center of purple bacteria, two homologous integral membrane proteins, L(ight) and M(edium), are known to be essential to the light-mediated water-splitting process. In the photosystem II of eukaryotic chloroplasts two related proteins are involved: the D1 (psbA) and D2 proteins (psbD). These four types of protein probably evolved from a common ancestor [see 1,2 for recent reviews].

A signature pattern was developed which include two conserved histidine residues. In L and M chains, the first histidine is a ligand of the magnesium ion of the special pair bacteriochlorophyll, the second is a ligand of a ferrous non-heme iron atom. In photosystem II these two histidines are thought to play a similar role.

Consensus pattern[NQH]-x(4)-P-x-H-x(2)-[SAG]-x(11)-[SAGC]-x-H-[SAG](2) [The first H is a magnesium ligand] [The second H is a iron ligand] Sequences known to belong to this class detected by the patternALL, except

782

for broad bean psbA which has Gln instead of the second His.

[1] Michel H., Deisenhofer J. Biochemistry 27:1-7(1988).

[2]Barber J. Trends Biochem. Sci. 12:321-326(1987).

5

987. phytochrome: Phytochrome region

This family contains a region specific to phytochrome proteins. Number of members:

145

10 988. PI3K_C2: C2 domain

Phosphoinositide 3-kinase region postulated to contain a C2 domain. Outlier of C2 family.

Number of members:

39

[1] Medline: 97388296 Using structure to define the function of phosphoinositide 3-kinase family members. Domin J, Waterfield MD; FEBS Lett 1997;410:91-95.

[2] Medline: 97398940 Phosphoinositide 3-kinases: a conserved family of signal transducers. Vanhaesebroeck B, Leevers SJ, Panayotou G, Waterfield MD; Trends Biochem Sci 1997;22:267-272.

989. PI3Ka: Phosphoinositide 3-kinase family, accessory domain (PIK domain)
PIK domain is conserved in all PI3 and PI4-kinases. Its role is unclear but it has been suggested [2] to be involved in substrate presentation.

Number of members: 47

- [1] Medline: 97388296 Using structure to define the function of phosphoinositide 3-kinase family members. Domin J, Waterfield MD; FEBS Lett 1997;410:91-95.
 - [2] Medline: 94069320 Phosphatidylinositol 4-kinase: gene structure and requirement for yeast cell viability. Flanagan CA, Schnieders EA, Emerick AW, Kunisawa R, Admon A, Thorner J; Science 1993;262:1444-1448.

30

990. P-II protein signatures

PROSITE cross-reference(s): PS00496; PII_GLNB UMP, PS00638; PII GLNB CTER

30

The P-II protein (gene glnB) is a bacterial protein important for the control of glutamine synthetase [1,2,3]. In nitrogen-limiting conditions, when the ratio of glutamine to 2-ketoglutarate decreases, P-II is uridylylated on a tyrosine residue to form P-II-UMP. P-II-UMP allows the deadenylation of glutamine synthetase (GS), thus activating the enzyme.

- Conversely, in nitrogen excess, P-II-UMP is deuridylated and then promotes the adenylation of GS. P-II also indirectly controls the transcription of the GS gene (glnA) by preventing NR-II (ntrB) to phosphorylate NR-I (ntrC) which is the transcriptional activator of glnA. Once P-II is uridylylated, these events are reversed.
- P-II is a protein of about 110 amino acid residues extremely well conserved. The tyrosine which is urydylated is located in the central part of the protein.

In cyanobacteria, P-II seems to be phosphorylated on a serine residue rather than being urydylated.

In methanogenic archaebacteria, the nitrogenase iron protein gene (nifH) is followed by two open reading frames highly similar to the eubacterial P-II protein [4]. These proteins could be involved in the regulation of nitrogen fixation.

In the red alga, Porphyra purpurea, there is a glnB homolog encoded in the chloroplast genome.

Other proteins highly similar to glnB are:

- Bacillus subtilis protein nrgB [5].
 - Escherichia coli hypothetical protein ybaI [6].

Two signature patterns were developed for P-II protein. The first one is a conserved stretch (in eubacteria) of six residues which contains the urydylated tyrosine, the other is derived from a conserved region in the C-terminal part of the P-II protein.

Consensus patternY-[KR]-G-[AS]-[AE]-Y [The second Y is uridylated] Sequences known to belong to this class detected by the patternALL glnB's from eubacteria.

Consensus pattern[ST]-x(3)-G-[DY]-G-[KR]-[IV]-[FW]-[LIVM]-x(2)-[LIVM]

[1]Magasanik B. Biochimie 71:1005-1012(1989).

[2]Holtel A., Merrick M. Mol. Gen. Genet. 215:134-138(1988).

5 [3]Cheah E., Carr P.D., Suffolk P.M., Vasuvedan S.G., Dixon N.E., Ollis D.L. Structure 2:981-990(1994).

[4] Sibold L., Henriquet M., Possot O., Aubert J.-P. Res. Microbiol. 142:5-12(1991).

[5] Wray L.V. Jr., Atkinson M.R., Fisher S.H. J. Bacteriol. 176:108-114(1994).

[6] Allikmets R., Gerrard B.C., Court D., Dean M.C. Gene 136:231-236(1993).

10

15

20

991. PIP5K: Phosphatidylinositol-4-phosphate 5-Kinase

This family contains a region from the common kinase core found in the type I phosphatidylinositol-4-phosphate 5-kinase (PIP5K) family as described in [1]. The family consists of various type I, II and III PIP5K enzymes. PIP5K catalyses the formation of phosphoinositol-4,5-bisphosphate via the phosphorylation of phosphatidylinositol-4-phosphate a precursor in the phosphinositide signaling pathway. Number of members:

33

- [1] Medline: 98204859. Type I phosphatidylinositol-4-phosphate 5-kinases. Cloning of the third isoform and deletion/substitution analysis of members of this novel lipid kinase family. Ishihara H, Shibasaki Y, Kizuki N, Wada T, Yazaki Y, Asano T, Oka Y; J Biol Chem 1998;273:8741-8748.
- [2] Medline: 97115834 Type I phosphatidylinositol-4-phosphate 5-kinases are distinct members of this novel lipid kinase family. Loijens JC, Anderson RA; J Biol Chem 1996 20:271:32937-32943.

25

992. PolyA pol: Poly A polymerase family

This family includes nucleic acid independent RNA polymerases, such as Poly(A) polymerase, which adds the poly (A) tail to mRNA EC:2.7.7.19. This family also includes the tRNA nucleotidyltransferase that adds the CCA to the 3' of the tRNA

30 EC:2.7.7.25. Number of members: 31

[1] Medline: 93066242 Identification of the gene for an Escherichia coli poly(A) polymerase. Cao GJ, Sarkar N; Proc Natl Acad Sci U S A 1992;89:10380-10384.

785

993. Photosystem I psaA and psaB proteins signature (psaA_psaB)
PROSITE cross-reference(s)PS00419; PHOTOSYSTEM I PSAAB

Photosystem I (PSI) [1] is an integral membrane protein complex that uses light energy to mediate electron transfer from plastocyanin to ferredoxin. PSI is found in the chloroplast of plants and cyanobacteria. The electron transfer components of the reaction center of PSI are a primary electron donor P-700 (chlorophyll dimer) and five electron acceptors: A0 (chlorophyll), A1 (a phylloquinone) and three 4Fe-4S iron-sulfur centers: Fx, Fa, and Fb.

PsaA and psaB, two closely related proteins, are involved in the binding of P700, A0, A1, and Fx. psaA and psaB are both integral membrane proteins of 730 to 750 amino acids that seem to contain 11 transmembrane segments. The Fx 4Fe-4S iron-sulfur center is bound by four cysteines; two of these cysteines are provided by the psaA protein and the two others by psaB. The two cysteines in both proteins are proximal and located in a loop between the ninth and tenth transmembrane segments. A leucine zipper motif seems to be present [2] downstream of the cysteines and could contribute to dimerization of psaA/psaB.

The signature pattern for these proteins is based on the perfectly conserved region that includes the two iron-sulfur binding cysteines.

20 Consensus patternC-D-G-P-G-R-G-G-T-C [The two C's bind the iron-sulfur center]

[1] Golbeck J.H. Biochim. Biophys. Acta 895:167-204(1987).[2] Webber A.N., Malkin R. FEBS Lett. 264:1-14(1990).

25 994. PSBH: Photosystem II 10 kDa phosphoprotein

This protein is phosphorylated in a light dependent reaction.

Number of members: 20

995. PsbJ

This family consists of the photosystem II reaction center protein PsbJ from plants and Cyanobacteria. In Synechocystis sp. PCC 6803 PsbJ regulates the number of photosystem II centers in thylakoid membranes, it is a predicted 4kDa protein with one membrane spanning domain [1]. Number of members: 20

[1] Medline: 93131892. Genetic and immunological analyses of the cyanobacterium Synechocystis sp. PCC 6803 show that the protein encoded by the psbJ gene regulates the number of photosystem II centers in thylakoid membranes. Lind LK, Shukla VK, Nyhus KJ, Pakrasi HB; J Biol Chem 1993;268:1575-1579.

5

10

15

20

25

30

996. PSBT: Photosystem II reaction centre T protein

The exact function of this protein is unknown. It probably consists of a single transmembrane spanning helix. The Swiss:P37256 protein, appears to be (i) a novel photosystem II subunit and (ii) required for maintaining optimal photosystem II activity under adverse growth conditions [1]. Number of members: 17

[1] Medline: 94298765. The chloroplast ycf8 open reading frame encodes a photosystem II polypeptide which maintains photosynthetic activity under adverse growth conditions. Monod C, Takahashi Y, Goldschmidt-Clermont M, Rochaix JD; EMBO J 1994;13:2747-2754.

997. PSI_8. PHOTOSYSTEM I REACTION CENTRE SUBUNIT VIII. Synonym(s)PSI-I. Gene name(s)PSAI. From Hordeum vulgare (Barley). Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.

MAY HELP IN THE ORGANIZATION OF THE PSAL SUBUNIT. BELONGS TO THE PSAI FAMILY.

[1] SEQUENCE FROM N.A. MEDLINE; 90036933. Scheller H.V., Okkels J.S., Hoej P.B., Svendsen I., Roepstorff P., Moeller B.L.; "The primary structure of a 4.0-kDa photosystem I polypeptide encoded by the chloroplast psaI gene."; J. Biol. Chem. 264:18402-18406(1989).

998. PSI_PsaJ: Photosystem I reaction centre subunit IX / PsaJ

This family consists of the photosystem I reaction centre subunit IX or PsaJ from various organisms including Synechocystis sp. (strain pcc 6803), Pinus thunbergii (green pine) and Zea mays (maize). PsaJ Swiss:P19443 is a small 4.4kDa, chloroplastal encoded, hydrophobic subunit of the photosystem I reaction complex its function is not yet fully understood [1].

PsaJ can be cross-linked to PsaF Swiss:P12356 and has a single predicted transmembrane

10

15

20

25

30

domain it has a proposed role in maintaing PsaF in the correct orientation to allow for fast electron transfer from soluble donor proteins to P700+ [1]. Number of members: 18

[1] Medline: 99238330. A large fraction of PsaF is nonfunctional in photosystem I complexes lacking the PsaJ subunit. Fischer N, Boudreau E, Hippler M, Drepper F, Haehnel W, Rochaix JD; Biochemistry 1999;38:5546-5552.

[2] Medline: 93252282. Genes encoding eleven subunits of photosystem I from the thermophilic cyanobacterium Synechococcus sp. Muhlenhoff U, Haehnel W, Witt H, Herrmann RG; Gene 1993;127:71-78.

999. PSII. Protein namePHOTOSYSTEM II P680 CHLOROPHYLL A APOPROTEIN. Synonym(s)CP-47 PROTEIN. Gene name(s)PSBB. From Hordeum vulgare (Barley), Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.

FUNCTION: THIS PROTEIN CONJUGATES WITH CHLOROPHYLL & CATALYZES THE PRIMARY LIGHT-INDUCED PHOTOCHEMICAL PROCESSES OF PHOTOSYSTEM II. SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE. SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.

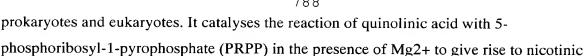
[1] SEQUENCE FROM N.A. STRAIN=CV. SABARLIS; MEDLINE; 89240047. Andreeva A.V., Buryakova A.A., Reverdatto S.V., Chakhmakhcheva O.G., Efimov V.A.; "Nucleotide sequence of the 5.2 kbp barley chloroplast DNA fragment, containing psbB-psbH-petB-petD gene cluster."; Nucleic Acids Res. 17:2859-2860(1989).

[2] SEQUENCE FROM N.A. STRAIN=CV. SABARLIS; MEDLINE; 92207253. Efimov V.A., Andreeva A.V., Reverdatto S.V., Chakhmakhcheva O.G.; "Photosystem II of rye. Nucleotide sequence of the psbB, psbC, psbE, psbF, psbH genes of rye and chloroplast DNA regions adjacent to them."; Bioorg. Khim. 17:1369-1385(1991).

[3] SEQUENCE OF 411-420. Hinz U.G.; "Isolation of the photosystem II reaction center complex from barley. Characterization by cicular dichroism spectroscopy and amino acid sequencing."; Carlsberg Res. Commun. 50:285-298(1985).

 $1000.\ QRPT ase.\ Quino linate\ phosphoribosyl\ transferase.$

Quinolinate phosphoribosyl transferase (QPRTase) or nicotinate-nucleotide pyrophosphorylase EC:2.4.2.19 is involved in the de novo synthesis of NAD in both



acid mononucleotide (NaMN), pyrophosphate and carbon dioxide [1,2]. Number of members:

26.

5

10

15

20

[1] Medline: 97169443. A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. Eads JC, Ozturk D, Wexler TB, Grubmever C, Sacchettini JC; Structure 1997;5:47-58.

[2] Medline: 96139309. The sequencing expression, purification, and steady-state kinetic analysis of quinolinate phosphoribosyl transferase from Escherichia coli. Bhatia R, Calvo KC; Arch Biochem Biophys 1996;325:270-278.

1001. R3H domain

The name of the R3H domain comes from the characteristic spacing of the most conserved arginine and histidine residues. The function of the domain is predicted to be binding ssDNA. Number of members: 28

[1] Medline: 99003905 The R3H motif: a domain that binds single-stranded nucleic acids. Grishin NV; Trends Biochem Sci 1998;23:329-330.

1002. recF protein signatures (RecF)

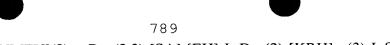
The prokaryotic protein recF [1,2] is a single-stranded DNA-binding protein which also probably binds ATP. RecF is involved in DNA metabolism; it is required for recombinational DNA repair and for induction of the SOS response. RecF is a protein of about 350 to 370 amino acid residues; there is a conserved ATP-binding site motif 'A' (P-loop) in the Nterminal section of the protein as well as two other conserved regions, one located in the central section, and the other in the C-terminal section. Signature patterns were derived from these two regions.

30

25

Consensus pattern [LIVM]-x(4)-[LIF]-x(6)-[LIF]-[LVF]-x-[GE]-[GSTAD]-[PA]-x(2)-R-Rx-[FYW]-[LIVMF]-D Sequences known to belong to this class detected by the pattern ALL.

30



Consensus pattern[LIVMFY](2)-x-D-x(2,3)-[SA]-[EH]-L-D-x(2)-[KRH]-x(3)-L Sequences known to belong to this class detected by the patternALL, except for T. palidum recF.

- [1] Sandler S.J., Chackerian B., Li J.T., Clark A.J. Nucleic Acids Res. 20:839-845(1992).
- 5 [2] Alonso J.C., Fisher L.M.; Mol. Gen. Genet. 246:680-686(1995).
 - 1003. RibD C-terminal domain (RibD_C)
- The function of this domain is not known, but it is thought to be involved in riboflavin biosynthesis. This domain is found in the C terminus of RibD/RibG Swiss:P25539, in combination with dCMP_cyt_deam, as well as in isolation in some archaebacterial proteins Swiss:P95872.

Number of members: 21

15 1004. Ribosomal protein L16 signatures (Ribosomal_L16)

Ribosomal protein L16 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L16 is known to bind directly the 23S rRNA and to be located at the A site of the peptidyltransferase center. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

- Eubacterial L16.
- Algal and plant chloroplast L16.
- Cyanelle L16.
- Plant mitochondrial L16.
- L16 is a protein of 133 to 185 amino-acid residues. As signature patterns, we selected two conserved regions in the central section of these proteins.

Consensus pattern [KR](2)-x-[GSAC]-[KRQVA]-[LIVM]-W-[LIVM]-[KR]-[LIVM]-[LFY]-[AP] Sequences known to belong to this class detected by the pattern ALL.

- Consensus patternR-M-G-x-[GR]-K-G-x(4)-[FWKR] Sequences known to belong to this class detected by the patternALL.
- [1] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

5

1005. Ribosomal protein L32e signature (Ribosomal L32E)

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L32 [1].
- Drosophila RP49 [2].
- Trichoderma harzianum L32 [3].
- Yeast L32e (YBL092w).
- Archaebacterial L32e [4].

These proteins have 135 to 240 amino-acid residues. As a signature pattern, a stretch of about 20 residues located in the N-terminal part of these proteins was selected.

Consensus patternF-x-R-x(4)-[KR]-x(2)-[KR]-[LIVMF]-x(3,5)-W-R-[KR]-x(2)-G Sequences known to belong to this class detected by the pattern ALL.

- [1] Jacks C.M., Powaser C.B., Hackett P.B. Gene 74:565-570(1988).
- [2] Aguade M. Mol. Biol. Evol. 5:433-441(1988).
- [3] Lora J.M., Garcia I., Benitez T., Llobell A., Pintor-Toro J.A. Nucleic Acids Res.
- 20 21:3319-3319(1993).
 - [4] Arndt E., Scholzen T., Kroemer W., Hatakeyama T., Kimura M. Biochimie 73:657-668(1991).

1006. (Ribosomal S3) Ribosomal protein S3 signature

PROSITE: PDOC00474. PROSITE cross-reference(s) PS00548; RIBOSOMAL_S3
Ribosomal protein S3 is one of the proteins from the small ribosomal subunit.

In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

- 30 -Eubacterial S3.
 - -Algal and plant chloroplast S3.
 - -Cyanelle S3.
 - -Archaebacterial S3.
 - -Plant mitochondrial S3.

15

20

25

30

- -Vertebrate S3.
- -Insect S3.
- -Caenorhabditis elegans S3 (C23G10.3).
- -Yeast S3 (Rp13).
- S3 is a protein of 209 to 559 amino-acid residues. A conserved region located in the C-terminal section was selected as a signature pattern.

Consensus pattern[GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS]. Sequences known to belong to this class detected by the patternALL, except for some mitochondrial S3.

[1]Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

1007. RimM - RimM

The RimM protein is essential for efficient processing of 16S rRNA [1]. The RimM protein was shown to have affinity for free ribosomal 30S subunits but not for 30S subunits in the 70S ribosomes [1]. Number of members: 14.

[1]Medline: 98083058. RimM and RbfA are essential for efficient processing of 16S rRNA in Escherichia coli. Bylund GO, Wipemo LC, Lundberg LA, Wikstrom PM; J Bacteriol 1998;180:73-82.

1008. RNA_pol_A - RNA polymerase alpha subunit

- -!- RNA polymerases catalyse the DNA dependent polymerisation of RNA. Prokaryotes contain a single RNA polymerase compared to three in eukaryotes (not including mitochondrial and chloroplast polymerases).
- -!- Members of this family include: A subunit from eukaryotes, gamma subunit from cyanobacteria, beta' subunit from eubacteria, A' subunit from archaebacteria, B" from chloroplasts. Number of members: 139.

[1]Medline: 97066998. Structural modules of the large subunits of RNA polymerase. Introducing archaebacterial and chloroplast split sites in the beta and beta' subunits of Escherichia coli RNA polymerase. Severinov K, Mustaev A, Kukarin A, Muzzin O, Bass I, Darst SA, Goldfarb A; J Biol Chem 1996;271:27969-27974.

10

15

20

1009. RuBisCO_large - Ribulose bisphosphate carboxylase large chain active site PROSITE: PDOC00142; PROSITE cross-reference(s) PS00157; RUBISCO_LARGE

Ribulose bisphosphate carboxylase (EC 4.1.1.39) (RuBisCO) [1,2] catalyzes the initial step in Calvin's reductive pentose phosphate cycle in plants as well as purple and green bacteria. It consists of a large catalytic unit and a small subunit of undetermined function. In plants, the large subunit is coded by the chloroplastic genome while the small subunit is encoded in the nuclear genome. Molecular activation of RuBisCO by CO2 involves the formation of a carbamate with the epsilon-amino group of a conserved lysine residue. This carbamate is stabilized by a magnesium ion. One of the ligands of the magnesium ion is an aspartic acid residue close to the active site lysine [3]. A pattern was developed which includes both the active site residue and the metal ligand, and which is specific to RuBisCO large chains.

Consensus patternG-x-[DN]-F-x-K-x-D-E [K is the active site residue] [The second D is a magnesium ligand]. Sequences known to belong to this class detected by the patternALL, except for Cheilopleuria biscuspis RuBisCO.

[1] Miziorko H.M., Lorimer G.H. Annu. Rev. Biochem. 52:507-535(1983).

[2] Akazawa T., Takabe T., Kobayashi H. Trends Biochem. Sci. 9:380-383(1984).[3] Andersson I., Knight S., Schneider G., Lindqvist Y., Lundqvist T., Branden C.-I., Lorimer G.H. Nature 337:229-234(1989).

1010. Rve - Integrase core domain

Integrase mediates integration of a DNA copy of the viral genome into the host chromosome. Integrase is composed of three domains. The amino-terminal domain is a zinc binding domain Integrase_Zn. This domain is the central catalytic domain. The carboxyl terminal domain that is a non-specific DNA binding domain integrase. The catalytic domain acts as an endonuclease when two nucleotides are removed from the 3' ends of the blunt-ended viral DNA made by reverse transcription. This domain also catalyses the DNA strand transfer reaction of the 3' ends of the viral DNA to the 5' ends of the integration site [1]. Number of members: 694.

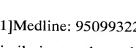
10

15

20

25

30



[1]Medline: 95099322. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR; Science 1994;266:1981-1986.

1011. (SBP_bac_3) Bacterial extracellular solute-binding proteins, family 3 signature PROSITE: PDOC00798. PROSITE cross-reference(s) PS01039; SBP_BACTERIAL_3

Bacterial high affinity transport systems are involved in active transport of solutes across the cytoplasmic membrane. The protein components of these traffic systems include one or two transmembrane protein components, one or two membrane-associated ATPbinding proteins (ABC transporters; see <PDOC00185>) and a high affinity periplasmic solute-binding protein. The later are thought to bind the substrate in the vicinity of the inner membrane, and to transfer it to a complex of inner membrane proteins for concentration into the cytoplasm.

In gram-positive bacteria which are surrounded by a single membrane and have therefore no periplasmic region the equivalent proteins are bound to the membrane via an Nterminal lipid anchor. These homolog proteins do not play an integral role in the transport process per se, but probably serve as receptors to trigger or initiate translocation of the solute throught the membrane by binding to external sites of the integral membrane proteins of the efflux system.

In addition at least some solute-binding proteins function in the initiation of sensory transduction pathways.

On the basis of sequence similarities, the vast majority of these solute-binding proteins can be grouped [1] into eight families of clusters, which generally correlate with the nature of the solute bound.

Family 3 groups together specific amino acids and opine-binding periplasmic proteins and a periplasmic homolog with catalytic activity:

- -Histidine-binding protein (gene hisJ) of Escherichia coli and related bacteria. An homologous lipoprotein exists in Neisseria gonorrhoeae.
- -Lysine/arginine/ornithine-binding proteins (LAO) (gene argT) of Escherichia coli and related bacteria are involved in the same transport system than his J. Both solute-binding proteins interact with a common membrane-bound receptor hisP of the binding protein dependent transport system HisQMP.
- -Glutamine-binding proteins (gene glnH) of Escherichia coli and Bacillus stearothermophilus.

30

- -Glutamate-binding protein (gene gluB) of Corynebacterium glutamicum.
- -Arginine-binding proteins artI and artJ of Escherichia coli.
- -Nopaline-binding protein (gene nocT) from Agrobacterium tumefaciens.
- -Octopine-binding protein (gene occT) from Agrobacterium tumefaciens.
- 5 -Major cell-binding factor (CBF1) (gene: peb1A) from Campylobacter jejuni.
 - -Bacteroides nodosus protein aabA.
 - -Cyclohexadienyl/arogenate dehydratase of Pseudomonas aeruginosa, a periplasmic enzyme which forms an alternative pathway for phenylalanine biosynthesis.
 - -Escherichia coli protein fliY.
- 10 -Vibrio harveyi protein patH.
 - -Escherichia coli hypothetical protein ydhW.
 - -Bacillus subtilis hypothetical protein yckB.
 - -Bacillus subtilis hypothetical protein yckK.
- 15 The signature pattern is located near the N-terminus of the mature proteins.

Consensus patternG-[FYIL]-[DE]-[LIVMT]-[DE]-[LIVMF]-x(3)-[LIVMA]-[VAGC]-x(2)-[LIVMAGN]

Sequences known to belong to this class detected by the patternALL.

20 [1]Tam R., Saier M.H. Jr. Microbiol. Rev. 57:320-346(1993).

1012. Sec7 - Sec7 domain

The Sec7 domain is a guanine-nucleotide-exchange-factor (GEF) for the arf family [2]. Number of members: 32.

[1] Medline: 98169075. Structure of the Sec7 domain of the Arf exchange factor. ARNO. Cherfils J, Menetrey J, Mathieu M, Le Bras G, Robineau S, Beraud-Dufour S, Antonny B, Chardin P; Nature 1998;392:101-105.

[2] Medline: 97100951. A human exchange factor for ARF contains Sec7- and pleckstrinhomology domains. Chardin P, Paris S, Antonny B, Robineau S, Beraud-Dufour S, Jackson CL, Chabre M. Nature 1996;384:481-484.

1013. SecA_protein. SecA protein, amino terminal region

25

30

5

SecA protein binds to the plasma membrane where it interacts with proOmpA to support translocation of proOmpA through the membrane. SecA protein achieves this translocation, in association with SecY protein, in an ATP dependent manner. SecA possesses the ATPase activity. The carboxyl terminus has similarity with the helicase carboxyl terminus. See Ribosomal L5. Number of members: 45.

[1]Medline: 98309858. Amino-terminal region of SecA is involved in the function of SecG for protein translocation into Escherichia coli membrane vesicles. Mori H, Sugiyama H, Yamanaka M, Sato K, Tagaya M, Mizushima S; J Biochem (Tokyo) 1998;124:122-129.

[2]Medline: 89251629. SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of Escherichia coli. Lill R, Cunningham K, Brundage LA, Ito K, Oliver D, Wickner W; EMBO J 1989;8:961-966.

1014. Seedstore_2S - 2S seed storage family

Members of this family are composed of two chains (both included in the alignment), these are co-translated and later cleaved. The two chains are disulphide linked together. Number of members: 27.

[1]Medline: 97121264. 1H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein. Rico M, Bruix M, Gonzalez C, Monsalve RI, Rodriguez R; Biochemistry 1996;35:15672-15682.

1015. Smr - Smr domain

This family includes the Smr (Small MutS Related) proteins, and the C-terminal region of the MutS2 protein. It has been suggested that this domain interacts with the MutS1 Swiss:P23909 protein in the case of Smr proteins and with the N-terminal MutS related region of MutS2 Swiss:P94545 [1]. Number of members: 14.

[1]Medline: 10431172. Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS2 family. Moreira D, Philippe H; Trends Biochem Sci 1999;24:298-300.

1016. (SSF) Sodium:solute symporter family signatures and profile PROSITE: PDOC00429. PROSITE cross-reference(s)PS00456; NA_SOLUT_SYMP_1 PS00457; NA_SOLUT_SYMP_2 PS50283; NA_SOLUTE_SYMP_3

5

796

It has been shown [1,2] that integral membrane proteins that mediate the intake of a wide variety of molecules with the concomitant uptake of sodium ions (sodium symporters) can be grouped, on the basis of sequence and functional similarities into a number of distinct families. One of these families is known as the sodium:solute symporter family (SSF) and currently consists of the following proteins:

- -Mammalian Na+/glucose co-transporter.
- -Mammalian Na+/myo-inositol co-transporter.
- -Mammalian Na+/nucleoside co-transporter.
- -Mammalian Na+/neutral amino acid co-transporter.
- -Escherichia coli Na+/proline symporter (gene putP).
 - -Escherichia coli Na+/pantothenate symporter (gene panF).
 - -Escherichia coli hypothetical protein yidK.
 - -Escherichia coli hypothetical protein yjcG.
 - -Bacillus subtilis hypothetical protein ywcA (ipa-31R).

These integral membrane proteins are predicted to comprise at least ten membrane spanning domains. Two conserved regions were selected as signature patterns; the first one is located in the fourth transmembrane region and the second one in a loop between two transmembrane regions in the C-terminal part of these proteins.

- Consensus pattern[GS]-x(2)-[LIY]-x(3)-[LIVMFYWSTAG](10)-[LIY]-[TAV]-x(2)-G-G-[LMF]-x-[SAP]. Sequences known to belong to this class detected by the patternALL. Consensus pattern[GAST]-[LIVM]-x(3)-[KR]-x(4)-G-A-x(2)-[GAS]-[LIVMGS]-[LIVMW]-[LIVMGAT]-G-x-[LIVMGA] Sequences known to belong to this class detected by the patternALL, except for E.coli yidK.
- Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.
 - [1] Reizer J., Reizer A., Saier M.H. Jr. Res. Microbiol. 141:1069-1072(1991).
- 30 [2]Reizer J., Reizer A., Saier M.H. Jr. Biochim. Biophys. Acta 1197:133-136(1994).

1017. SurE - Survival protein SurE

E. coli cells with the surE gene disrupted are found to survive poorly in stationary phase [1]. It is suggested that SurE may be involved in stress response. Yeast also contains a member of

10

15

20

25

797

the family Swiss:P38254. Swiss:P30887 can complement a mutation in acid phosphatase, suggesting that members of this family could be phosphatases. Number of members: 17.

[1]Medline: 95014035. A new gene involved in stationary-phase survival located at 59 minutes on the Escherichia coli chromosome. Li C, Ichikawa JK, Ravetto JJ, Kuo HC, Fu JC, Clarke S; J Bacteriol 1994;176:6015-6022.

[2]Medline: 93046805. Complementation of Saccharomyces cerevisiae acid phosphatase mutation by a genomic sequence from the yeast Yarrowia lipolytica identifies a new phosphatase. Treton BY, Le Dall MT, Gaillardin CM; Curr Genet 1992;22:345-355.

1018. Synuclein - Synuclein

There are three types of synucleins in humans, these are called alpha, beta and gamma. Alpha synuclein has been found mutated in families with autosomal dominant Parkinson's disease. A peptide of alpha synuclein has also been found in amyloid plaques in Alzheimer's patients. Number of members: 12.

[1]Medline: 98424410. The synuclein family. Lavedan C; Genome Res 1998;8:871-880.

1019. (T-box) T-box domain signatures

PROSITE: PDOC00972. PROSITE cross-reference(s) PS01283; TBOX_1 PS01264; TBOX_2

A number of eukaryotic DNA-binding proteins contain a domain of about 170 to 190 amino acids known as the T-box domain [1,2,3] and which probably binds DNA. The T-box has first been found in the mice T locus (Brachyury) protein, a transcription factor involved in mesoderm differentiation. It has since been found in the following proteins:

- -Vertebrate and invertebrate homologs of the T protein.
- -Mammalian proteins TBX1 to TBX6.
- -Mammalian protein TBR1 which is expressed specifically in brain.
- -Xenopus laevis eomesodermin (eomes).
- -Xenopus laevis Vegt (or Antipodean), a transcription factor that activates the expression of wnt-8, eomes and Brachyury.
 - -Chicken TbxT.
 - -Drosophila protein optomotor-blind (omb).
 - -Drosophila protein brachyenteron (byn) (also known as Trg), which is

5



required for the specification of the hindgut and anal pads.

- -Drosophila protein H15.
- -Caenorhabditis elegans protein tbx-12.
- -Caenorhabditis elegans hypothetical proteins F21H11.3, F40H6.4, T07C4.2, T07C4.6 and ZK177.10.

798

Two conserved regions were selected as signature patterns for the T-domain. The first region corresponds to the N-terminal of the domain and the second one to the central part.

Consensus patternL-W-x(2)-[FC]-x(3,4)-[NT]-E-M-[LIV](2)-T-x(2)-G-[RG]-[KRQ]

Sequences known to belong to this class detected by the patternALL, except for C.elegans ZK177.10.

Consensus pattern[LIVMYW]-H-[PADH]-[DEN]-[GS]-x(3)-G-x(2)-W-M-x(3)-[IVA]-x- F Sequences known to belong to this class detected by the patternALL, except for C.elegans tbx-12, ZK177.10 and Drosophila H15.

[1]Bollag R.J., Siegfried Z., Cebra-Thomas J.A., Garvey N., Davison E.M., Silver L.M. Nat. Genet. 7:383-389(1994).

- [2] Agulnik S.I., Garvey N., Hancock S., Ruvinsky I., Chapman D.L., Agulnik I., Bollag R.J., Papaioannou V.E., Silver L.M. Genetics 144:249-254(1996).
- 20 [3]Papaioannou V.E. Trends Genet. 13:212-213(1997).

1020. Toprim - Toprim domain

This is a conserved region from DNA primase. This corresponds to the Toprim domain common to DnaG primases, topoisomerases, OLD family nucleases and RecR proteins [1].

- Both DnaG motifs IV and V are present in the alignment, the DxD (V) motif may be involved in Mg2+ binding and mutations to the conserved glutamate (IV) completely abolish DnaG type primase activity [1]. DNA primase EC:2.7.7.6 is a nucleotidyltransferase it synthesizes the oligoribonucleotide primers required for DNA replication on the lagging strand of the replication fork; it can also prime the leading stand and has been implicated in cell division [2]. Number of members: 133.
 - [1]Medline: 98391745. Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Aravind L, Leipe DD, Koonin EV; Nucleic Acids Res 1998;26:4205-4213.

25

5

10



[2]Medline: 97368180. Cloning and analysis of the dnaG gene encoding Pseudomonas putida DNA primase. Szafranski P, Smith CL, Cantor CR; Biochim Biophys Acta 1997;1352:243-248.

[3]Medline: 94124015. The Haemophilus influenzae dnaG sequence and conserved bacterial primase motifs. Versalovic J, Lupski JR; Gene 1993;136:281-286.

1021. TraB - TraB family

pAD1 is a hemolysin/bacteriocin plasmid originally identified in Enterococcus faecalis DS16. It encodes a mating response to a peptide sex pheromone, cAD1, secreted by recipient bacteria. Once the plasmid pAD1 is acquired, production of the pheromone ceases--a trait related in part to a determinant designated traB. However a related protein is found in C. elegans Swiss:Q94217, suggesting that members of the TraB family have some more general function. Number of members: 12.

[1]Medline: 94302142. Characterization of the determinant (traB) encoding sex pheromone shutdown by the hemolysin/bacteriocin plasmid pAD1 in Enterococcus faecalis. An FY, Clewell DB; Plasmid 1994;31:215-221.

1022. (Transpo_mutator) Transposases, Mutator family, signature PROSITE: PDOC00770. PROSITE cross-reference(s) PS01007; TRANSPOSASE MUTATOR

Autonomous mobile genetic elements such as transposon or insertion sequences (IS) encode an enzyme, called transposase, required for excising and inserting the mobile element. On the basis of sequence similarities, transposases can be grouped into various families. One of these families has been shown [1,2,3,E1] to consist of transposases from the following elements:

- -Mutator from Maize.
- -Is1201 from Lactobacillus helveticus.
- -Is905 from Lactococcus lactis.
- 30 -Is1081 from Mycobacterium bovis.
 - -Is6120 from Mycobacterium smegmatis.
 - -Is406 from Pseudomonas cepacia.
 - -IsRm3 from Rhizobium meliloti.
 - -IsRm5 from Rhizobium meliloti.

10

15

- -Is256 from Staphylococcus aureus.
- -IsT2 from Thiobacillus ferrooxidans.

The maize Mutator transposase (MudrA) is a protein of 823 amino acids; the bacterial transposases listed above are proteins of 300 to 420 amino acids. These proteins contain a conserved domain of about 130 residues; a signature pattern was derived from the most conserved part of this domain.

Consensus patternD-x(3)-G-[LIVMF]-x(6)-[STAV]-[LIVMFYW]-[PT]-x-[STAV]-x(2)-[QR]-x-C-x(2)-H. Sequences known to belong to this class detected by the patternALL.

[1]Eisen J.A., Benito M.-I., Walbot V. Nucleic Acids Res. 22:2634-2636(1994).
[2]Guilhot C., Gicquel B., Davies J., Martin C. Mol. Microbiol. 6:107-113(1992).
[3]Wood M.S., Byrne A., Lessie T.G. Gene 105:101-105(1991).

1023. Transposase 8 - Transposase

Transposase proteins are necessary for efficient DNA transposition. This family consists of various E. coli insertion elements and other bacterial transposases some of which are members of the IS3 family. Number of members: 58.

- [1]Medline: 97324595. Genetic organization and transposition properties of IS511. D. A. Mullin, D. L. Zies, A. H. Mullin, N. Caballera & B. Ely; Mol Gen Genet 1997;254:456-463.
 [2]Medline: 97128810. The use of an improved transposon mutagenesis system for DNA sequencing leads to the characterization of a new insertion sequence of Streptomyces lividans 66. J. Fischer, H. Maier, P. Viell & J. Altenbuchner; Gene 1996;180:81-89.
- [3]Medline: 97074647. Identification and nucleotide sequence of Rhizobium meliloti insertion sequence ISRm6, a small transposable element that belongs to the IS3 family. S. Zekri & N. Toro; Gene 1996;175:43-48.
 - 1024. tRNA_int_endo tRNA intron endonuclease
- Members of this family cleave pre tRNA at the 5' and 3' splice sites to release the intron EC:3.1.27.9. Number of members: 8.

15

20

30



[1]Medline: 97344075. Properties of H. volcanii tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. Kleman-Leyer K, Armbruster DW, Daniels CJ; Cell 1997;89:839-847.

5 1025. Urease - Urease signatures

PROSITE: PDOC00133PROSITE cross-reference(s) PS01120; UREASE_1 PS00145; UREASE_2

Urease (EC 3.5.1.5) is a nickel-binding enzyme that catalyzes the hydrolysis of urea to carbon dioxide and ammonia [1]. Historically, it was the first enzyme to be crystallized (in 1926). It is mainly found in plant seeds, microorganisms and invertebrates. In plants, urease is a hexamer of identical chains. In bacteria [2], it consists of either two or three different subunits (alpha, beta and gamma).

Urease binds two nickel ions per subunit; four histidine, an aspartate and a carbamated-lysine serve as ligands to these metals; an additional histidine is involved in the catalytic mechanism [3].

As signatures for this enzyme, a region that contains two histidine that bind one of the nickel ions and the region of the active site histidine was selected.

Consensus pattern T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P [The two H's bind nickel]. Sequences known to belong to this class detected by the patternALL. Consensus pattern[LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A [H is the active site residue]. Sequences known to belong to this class detected by the patternALL.

[1] Takishima K., Suga T., Mamiya G. Eur. J. Biochem. 175:151-165(1988).

[2] Mobley H.L.T., Husinger R.P. Microbiol. Rev. 53:85-108(1989).
[3] Jabri E., Carr M.B., Hausinger R.P., Karplus P.A. Science 268:998-1004(1995).

1026. Urease_beta - Urease beta subunit.

This subunit is known as alpha in Heliobacter. Number of members: 35.

[1]Medline: 95273988. The crystal structure of urease from Klebsiella aerogenes. Jabri E, Carr MB, Hausinger RP, Karplus PA; Science 1995;268:998-1004.

1027. UvrD-helicase - UvrD/REP helicase

20

25



The Rep family helicases are composed of four structural domains. The Rep family function as dimers. REP helicases catalyse ATP dependent unwinding of double stranded DNA to single stranded DNA. Swiss:P23478, Swiss:P08394 have large insertions near to the carboxyterminus relative to other members of the family. Number of members: 52.

5

- [1] Medline: 97433075. Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. Korolev S, Hsieh J, Gauss GH, Lohman TM, Waksman G; Cell 1997;90:635-647.
- 10 1028. V-type ATPase 116kDa subunit family (V ATPase sub a)

This family consists of the 116kDa V-type ATPase (vacuolar (H+)-ATPases) subunits, as well as V-type ATP synthase subunit i. The V-type ATPases family are proton pumps that acidify intracellular compartments in eukaryotic cells for example yeast central vacuoles, clathrin-coated and synaptic vesicles. They have important roles in membrane trafficking processes [1]. The 116kDa subunit (subunit a) in the V-type ATPase is part of the V0 functional domain responsible for proton transport. The a subunit is a transmembrane glycoprotein with multiple putative transmembrane helices t has a hydrophilic amino terminal and a hydrophobic carboxy terminal [1,2]. It has roles in proton transport and assembly of the V-type ATPase complex [1,2]. This subunit is encoded by two homologous gene in yeast VPH1 and STV1 [2].

Number of members: 27

- [1] Forgac M; Medline: 99240666 Structure and properties of the vacuolar (H+)-ATPases." J Biol Chem 1999;274:12951-12954.
 - [2] Forgac M; Medline: 99270697 Structure and properties of the clathrin-coated vesicle and yeast vacuolar V-ATPases." J Bioenerg Biomembr 1999;31:57-65.
 - 1029. Viral (Superfamily 1) RNA helicase (Viral helicase1)
- Number of members: 260 30
 - [1] Koonin EV, Dolja VV; Medline: 94094568 Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences." Crit Rev Biochem Mol Biol 1993;28:375-430.

10

15

20

30

1030. Vesicular monoamine transporter (VMAT)

This family consists of various vesicular amine transporters with 12 transmembrane helices. These included vesicular acetylcholine transporters (VAChT) [3], and vesicular monoamine transporters (VMATs) [1,2] isoforms 1 adrenal and 2 brain (VMAT1 and VMAT2).

These proteins transport biogenic amines into synaptic vesicles or chromaffin granules [4]. VMATs pack monoamine neurotransmitters into secretary vesicles for regulated exocytotic release, they also protect against the parkinsonian neurotoxins MPP+ by transporting it into vesicles preventing it from acting on mitochondria [1].

Also in the family is C. elegans UNC-17 a putative vesicular acetylcholine transporter mutations in UNC-17 cause impaired neuromuscular function, giving rise to jerky or uncoordinated movement, [4].

Number of members: 15

- [1] Krantz DE, Peter D, Liu Y, Edwards RH; Medline: 97197857 Phosphorylation of a vesicular monoamine transporter by casein kinase II." J Biol Chem 1997;272:6752-6759.
- [2] Erickson JD, Varoqui H, Schafer MK, Modi W, Diebler MF, Weihe E, Rand J, Eiden LE, Bonner TI, Usdin TB; Medline: 94350930 Functional identification of a vesicular acetylcholine transporter and its expression from a 'cholinergic' gene locus." J Biol Chem 1994;269:21929-21932.
- [3] Erickson JD, Schafer MK, Bonner TI, Eiden LE, Weihe E; Medline: 96209876 Distinct
 pharmacological properties and distribution in neurons and endocrine cells of two isoforms of the human vesicular monoamine transporter." Proc Natl Acad Sci U S A 1996;93:5166-5171.
 [4] Alfonso A, Grundahl K, Duerr JS, Han HP, Rand JB; Medline: 3342494 The Caenorhabditis elegans unc-17 gene: a putative vesicular acetylcholine transporter." Science 1993;261:617-619.

1031. WW/rsp5/WWP domain signature and profile. Cross-reference(s): PS01159; WW_DOMAIN_1; PS50020; WW_DOMAIN_2

10

15

20

30

The WW domain [1-4,E1] (also known as rsp5 or WWP) has been originally discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown [5] to bind proteins with particular proline-motifs, [AP]-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

Proteins containing the WW domain are listed below.

- --Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C-terminal globular domain. Dystrophin form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats.
- --Utrophin, a dystrophin-like protein of unknown function.
- --Vertebrate YAP protein is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains [6].
- 25 --Mouse NEDD-4 plays a role in the embryonic development and differentiation of the central nervous system. It contains 3 WW modules followed by a HECT domain. The human ortholog contains 4 WW domains, but the third WW domain is probably spliced resulting in an alternate NEDD-4 protein with only 3 WW modules [3].
 - --Yeast RSP5 is similar to NEDD-4 in its molecular organization. It contains an N-terminal C2 domain (see <PDOC00380>), followed by a histidine-rich region, 3 WW domains and a HECT domain.
 - --Rat FE65, a transcription-factor activator expressed preferentially in liver. The activator domain is located within the N-terminal 232 residues of FE65, which also contain the WW domain.

10



- --Yeast ESS1/PTF1, a putative peptidyl prolyl cis-trans isomerase from family ppiC (see <PDOC00840>). A related protein, dodo (gene dod) exists in Drosophila and in mammals (gene PIN1).
- --Tobacco DB10 protein. The WW domain is located N-terminal to the region with similarity to ATP-dependent RNA helicases.
- --IQGAP, a human GTPase activating protein acting on ras. It contains an N-terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain.
- --Yeast pre-mRNA processing protein PRP40, Caenorhabditis elegans ZK1098.1 and fission yeast SpAC13C5.02 are related proteins with similarity to MYO2-type myosin, each containing two WW-domains at the N-terminus.
- --Caenorhabditis elegans hypothetical protein C38D4.5, which contains one WW module, a PH domain (see <PDOC50003>) and a C-terminal phosphatidylinositol 3-kinase domain.
- --Yeast hypothetical protein YFL010c.
- For the sensitive detection of WW domains, a profile was developed which spans the whole homology region as well as a pattern.

Description of pattern(s) and/or profile(s):

- Consensus patternW-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.
 - [1] Bork P., Sudol M. Trends Biochem. Sci. 19:531-533(1994).
 - [2] Andre B., Springael J.Y. Biochem. Biophys. Res. Commun. 205:1201-1205(1994).
 - [3] Hofmann K.O., Bucher P. FEBS Lett. 358:153-157(1995).
- 25 [4] Sudol M., Chen H.I., Bougeret C., Einbond A., Bork P. FEBS Lett. 369:67-71(1995).
 - [5] Chen H.I., Sudol M. Proc. Natl. Acad. Sci. U.S.A. 92:7819-7823(1995).
 - [6] Sudol M., Bork P., Einbond A., Kastury K., Druck T., Negrini M., Huebner K., Lehman D. J. Biol. Chem. 270:14733-14741(1995).
- 1032. XPA protein signatures. cross-reference(s): XPA_1 PROSITE PS00752; PS00753;XPA 2.

Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due

to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. XP-A is the most severe form of the disease and is due to defects in a 30 Kd nuclear protein called XPA (or XPAC) [2].

5

The sequence of the XPA protein is conserved from higher eukaryotes [3] to yeast (gene RAD14) [4]. XPA is a hydrophilic protein of 247 to 296 amino-acid residues which has a C4-type zinc finger motif in its central section.

10

15

20

Two signature were developed patterns for XPA proteins. The first corresponds to the zinc finger region, the second to a highly conserved region located some 12 residues after the zinc finger region.

Consensus patternC-x-[DE]-C-x(3)-[LIVMF]-x(1,2)-D-x(2)-L-x(3)-F-x(4)-C-x(2)-C Consensus pattern[LIVM](2)-T-[KR]-T-E-x-K-x-[DE]-Y-[LIVMF](2)-x-D-x-[DE]

[1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994).

- [2] Miura N., Miyamoto I., Asahina H., Satokata I., Tanaka K., Okada Y. J. Biol. Chem. 266:19786-19789(1991).
- [3] Shimamoto T., Kohno K., Tanaka K., Okada Y. Biochem. Biophys. Res. Commun. 181:1231-1237(1991).
- [4] Bankmann M., Prakash L., Prakash S. Nature 355:555-558(1992).

1033. YCF9

25

This family consists of the hypothetical protein product of the YCF9 gene from chloroplasts and cyanobacteria. Number of members: 16

1034. (DUF15)

30 It is highly con

It is highly conserved between eubacteria and eukaryotes.

Number of members: 30

1035. Lumenal portion of Cytochrome b559, alpha (gene psbE) subunit. (cytochr_b559a)

This family is the lumenal portion of cytochrome b559 alpha chain, matches to this family should be accompanied by a match to the cytochr_b559 family also. The Prosite pattern pattern matches the transmembrane region of the cytochrome b559 alpha and beta subunits.

Number of members: 16

A. Asparaginase 2

10

15

5

Asparaginase II (L-asparagine aminohydrolase II) is an extracellular protein that may be associated with the cell wall and whose expression is affected by the availability of nitrogen. Asparaginase II catalyzes the reaction of L-Asparagine + H₂O = L-Asparate + NH₃. As many leukemias have high requirements for aspartic acid, asparaginase II proteins are useful as reagents for screening compounds for activity as leukemia chemotherapy products. Asparaginase II protein can also be over- or under-expressed to alter amino acid content in plant tissues or to modify nitrogen fixation and/or nitrogen metabolism in plants.

Ref: Bon et al. (1997) Appl Biochem Biotechnol 63-65: 203-12

20

25

30

B. Chloroa b-bind

Chlorophyll a-b binding proteins are located in the thylakoid membranes of the chloroplast and bind chlorophyll a and chlorophyll b, thereby triggering a chemical reaction (photosynthesis). These proteins are useful in controlling the rate, efficiency and/or output of photosynthesis. Overexpression of chlorophyll a-b binding proteins is expected to increase the rate of photosynthesis.

Ref:

Leutwiler et al. (1986) Nucleic Acids Res 14: 4051-64 Brandt et al. (1992) Plant Mol Biol 19: 699-703

C. DMRL synthase

10

15

20

25

30



DMRL Synthase (6,7-Dimethyl-8-Ribityllumazine Synthase) catalyzes the last step in riboflavin (Vitamin B₂) synthesis, condensing 5-amino-6-(1'-D)-ribityl-amino-2,4(1H, 3H)-Pyrimidinedione with L-3,4-Dihydroxy-2-Butanone 4-Phosphate producing 6,7-Dimethyl-8-(1-D-Ribityl)Luminazine. The enzyme forms a homopentamer. Engineering of these proteins or those with homologous sequences/structures may allow control of the amounts of vitamin B₂ available in plants and/or accumulation of pigment, as well as altering reactions requiring hydrogen ion carriers/transmitters.

Ref: Garcia-Ramirez et al. (1995) J Biol Chem 270: 23801-7

D. E1 N

These proteins are ATP-dependent DNA helicases that are required for initiation of viral DNA replication. They form a complex with the viral E2 protein. The E1-E2 complex binds to the replication origin that contains binding sites for both proteins. The majority of sequences known for this group of proteins are from various papillomaviruses, a type of double stranded DNA virus. In plants, the prototype double stranded DNA virus is Cauliflower Mosaic virus (CaMV). Manipulation of these proteins, especially to produce variant proteins that form non-productive complexes, enables production of plants that are resistant to infection by double stranded DNA viruses.

Ref: Yang et al. (1993) PNAS USA 90: 5086-90 Ustav and Stenlund (1991) EMBO J 10: 449-57 Callaway et al. (1996) Mol Plant Microbe Interact 9: 810-8

E. EF1 G

Elongation Factor-1 is composed of four subunits: alpha, beta, delta and gamma. Gamma subunits are presumed to play a role in anchoring the complex to other cellular components. Studies of EF-1 genes in plants suggests that different forms of the EF-1 subunits may be expressed in particular organs or in response to stress. Manipulation of the activity of these proteins, either by altered expression level or by structural mutation, may result in the accumulation of a particular protein in a chosen organ or allow production of particular proteins during stress conditions.

5

Ref:

Kinzy et al. (1994) NAR 22: 2703-7

Dunn et al. (1993) Plant Mol Biol 23: 221-5

Aguilar et al. (1991) Plant Mol Biol 17: 351-60

F. ENV polyprotein

This family comprises the envelope or coat proteins known from a number of different retroviruses. In mammalian species, retroviruses are responsible for diseases such as leukemia and HIV. In plants, retroviruses are known in both monocot (e.g. Zeon-1) and dicot (e.g. Arabidopsis and tobacco) species and have been shown to induce mutant alleles at new loci. Engineering of plant ENV proteins may allow mobilization or targeting of endogenous or introduced retroviruses, in essence generating a new method for mutant production, gene tagging and the like.

15

Desegge Loreo

Ref: Mamoun et al (1990) J Virol 64: 4180-8

Grandbastien et al. (1989) Nature 337: 376-80

Wright and Voytas (1998) Genetics 149: 703-15

20

25

G. Glycosyl hydr9

Proteins having this domain (previously known as the glycosyl hydrolase family 5 domain) catalyze the endohydrolysis of 1,4-β-D-glucosidic linkages in cellulose. Numerous plant proteins with this domain exist and are expressed in an organ specific manner. They are involved in the fruit ripening process, in cell elongation and plant reproduction. Modulation of the activity of these proteins, either by over- or under-expression or by mutation of the polypeptide, could be used to affect post-harvest physiology (e.g. rate of ripening) or for engineering reproductive sterility.

30

Ref: Giorda et al. (1990) Biochemistry 29: 7264-9

Tucker et al. (1988) Plant Physiol 88: 1257-62

Shani et al. (1997) 43: 837-42

10

15

20

25

30

Milligan and Gasser (1995) Plant Mol Biol 28: 691-711

H. Glycosyl hydr14

The β -amylases (family 14 of glycosyl hydrolases) catalyze the hydrolysis of 1,4- α -glucosidic linkages in polysaccharides and remove successive maltose units from the non-reducing ends of the chains. Mutants of β -amylase in Arabidopsis exhibited altered degradation of starch throughout the diurnal cycle. In addition, the mutant phenotypes indicated that these enzymes not only affect carbohydrate metabolism/catabolism, but also influence the amount of pigment stored within particular cells. Manipulation of the β -amylase genes enables control of plant pigmentation (for example, fibre pigment in cotton) as well as carbohydrate synthesis and degradation.

Ref: Zeeman et al. (1998) Plant J 15: 357-65

Hirano and Nakamura (1997) Plant Physiol 114: 5675-82

Kitamoto et al. (1988) J Bacteriol 170: 5848-54

I. Glycosyl hydr15

Glycosyl hydrolases from family 15 (such as 1,4-Alpha-D-Glucan glucohydrolase,) catalyze the hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from the non-reducing ends of the chains resulting in the release of β -D-Glucose. In plants these proteins have been tied to the mobilization of the xyloglucan stored in the cotyledonary cell walls. Proteins such as these could be varied to affect the rate of plant growth (for example during germination), storage and/or use of glucose and other sugars by plant tissues and alteration of the properties, such as elasticity, of plant cell walls.

Ref: Crombie et al. (1998) Plant J 15: 27-38 Hata et al. (1991) Agric Biol Chem 55: 941-9

J. Glycosyl hydr20

5

10

15

20

Members of the family 20 glycosyl hydrolases catalyze the hydrolysis of terminal non-reducing N-acetly-D-hexosamine residues in N-acetyl-β-D-hexosaminides. N-acetyl-β-glucosaminidase belongs to this family and exists in several different forms (consisting of various combinations of alpha and beta chains) depending on the organism. Family 20 glycosyl hydrolases have been implicated in lysosomal storage diseases (such as Sandhoff disease) and glycogen storage disease in humans. These types of proteins are also responsible for the hydrolysis of chitin. In plants, these proteins could be useful in controlling carbohydrate catabolism, thereby influencing the amount of sugars available for storage and/or use in other metabolic pathways. In addition, it is possible that such proteins could be used to engineer an endogenous insect protection mechanism, e.g. by secretion of a chitin-hydrolyzing composition by the plant.

Ref: Graham et al (1988) J Biol Chem 263: 16823-9 O'Dowd et al. (1988) Biochemistry 27: 5216-26

K. HMG box

The HMG box is a novel type of DNA-binding domain found in a diverse group of proteins. Numerous plant proteins contain this domain, such as the HMG1/2-like proteins. The expression of some of these HMG proteins appears to be regulated by circadian rhythms and in a light dependent manner, occurring at higher levels in roots, for example and lower levels in light-grown tissues such as cotyledons. Generally, HMG proteins are thought to influence transcription regulation. In plants, HMGs are believed to have a role in maintaining patterns of circadian-regulated expression for other genes, suggesting that these proteins could be exploited to control growth and development.

Ref: Laudet et al. (1993) Nucleic Acids Res 21: 2493-501
Zheng et al. (1993) Plant Mol Biol 23: 813-23
Grasser et al. (1993) Plant Mol Biol 23: 619-25

30

25

20

30

5

Interleukin-2 (IL-2)is produced in mammals by T cells in response to antigenic or mitogenic stimulation and is crucial for proper regulation and functioning of the immune response. IL-2 is capable of stimulating B cells, monocytes, lymphokine-activated killer cells, natural killer cells and glioma cells. Plant extracts have also been shown to stimulate the immune system (for example, mistletoe therapy for human cancer). It is known that IL-2 is involved in feedback inhibition pathways that impact the inflammatory response as well as the growth inhibition of tumor reactive T cells. Plant proteins containing IL-2-like sequences are useful as immunity-based therapeutics, acting in a manner similar to IL-2 in mammals.

10 Ref: Heike et al. (1997) Scand J Immunol 45: 221-6
Ariel et al. (1998) J Immunol 161: 2465-72
Schink (1997) Anticancer Drugs 8 Suppl 1: S47-51

M. Oxidored FMN

NADPH dehydrogenases catalyze the reaction NADPH + acceptor = NADP(+) + reduced acceptor. One member of this family is yeast old yellow enzyme" (OYE) and is thought to be involved in oxylipin metabolism. A second yeast family member is a protein that binds estrogen binding protein (EBP) in addition to exhibiting oxidoreductase activity. An Arabidopsis homolog to OYE has been described and estrogen binding proteins in plants have been reported. Plant proteins from this class have the potential to be used to modify lipid metabolism/catabolism. These proteins may also have use as therapeutics for breast and prostate cancer, and other abnormal growth in steroid-sensitive tissues.

Ref: Baker et al. (1998) Proc Soc Exp Biol Med 217: 317-21
 Schaller and Weiler (1997) J Biol Chem 272: 28066-72
 Mandani et al. (1994) PNAS USA 91: 922-6

N. Oxidored q2

The NADH-plastoquinone oxidoreductases catalyze the reaction NADH + plastoquinone = NAD(+) + plastoquinol. In plants these reactions occur in the chloroplast and are believed to participate in a chloroplast respiratory system. Here, the NDH complex is postulated to act as

25

30

5

a valve to remove excess reduction equivalents in the chloroplasts. Manipulation of these proteins may improve the rate or efficiency of photosynthesis.

Ref: Burrows et al. (1998) EMBO J 17: 868-76 Kofer et al (1998) Mol Gen Genet 258: 166-73 Maier et al. (1995) J Mol Biol 251: 614-28

O. PABP

Polyadenylate binding proteins bind the poly (A) tail of mRNA. Plants, as exemplified by Arabidopsis, contain numerous PABP genes that are expressed in an organ-specific manner. For example, PABP2 is functional in roots and shoots, while PABP5 is expressed predominantly in immature flowers. The PABP proteins are implicated in numerous aspects of posttranscriptional regulation including mRNA turnover and translational initiation.

Control of activity of PABP proteins provides the ability to control the expression of various genes in particular organs during development.

Ref: Hilson et al (1993) Plant Physiol 103: 525-33
Belostotsky and Meagher (1993) PNAS USA 90: 6686-90

P. Parvo coat

Parvoviruses are linear single-stranded DNA viruses that are encapsulated by three capsid proteins. Plants are susceptible to infection by single stranded DNA viruses such as Maize streak virus (MSV) and various Gemini viruses. The coat proteins in these plant viruses are critical to the virus life cycle within the plant. For example, the coat protein of MSV is thought to be involved in intra- and inter-cellular movement within the plant. Engineering of proteins having similarity to parvoviral coat proteins, especially to produce proteins that interfere with maturation of the virus particle, enables the production of plants having better resistance to natural plant single-stranded DNA viruses.

Ref: Liu et al. (1997) J Gen Virol 78: 1265-70 Rohde et al. (1990) Virology 176: 648-51

10

15

20

25

30

O. Pkinase C

Plant serine/threonine protein kinases possessing this domain are expressed in all tissues and are known to undergo serine-specific autophosphorylation and specifically phosphorylate two ribosomal proteins, P14 and P16. During development, these proteins predominate during high metabolic activity in growing buds, root tips, leaf margins and germinating seeds. They are thought to be involved in the control of plant growth and development. In addition, two genes encoding proteins from this family have been described that help plant cells adapt during cold or high salt stresses. Consequently, engineering Pkinase C proteins provides a way to control general growth/development of the plant as well as a means to provide endogenous protection against environmental stresses.

Ref: Zhang et al. (1994) J Biol Chem 269: 17586-92 Mizoguchi et al. (1995) FEBS Lett 358: 199-204

R. REV

The REV proteins act post-transcriptionally to relieve negative repression of GAG and ENV production in retroviruses such as Human Immounodeficiency Virus type I (HIV-1). Plants contain retrovirus-like viruses such as pararetroviruses and retrotransposons (i.e. transposons having long terminal repeats). Plant retrotransposons in particular have been used to create mutations at various loci, thereby permitting gene isolation, gene tagging and the like. Manipulation of plant REV proteins enables control of transposition frequencies of corresponding transposable elements and provides a new tool for genetic engineering of plants.

Ref: Sodroski et al. (1986) Nature 321: 412-7
Franchini et al. (1989) PNAS USA 86: 2433-7
Marquet et al. (1995) 77: 113-24
Grandbastien et al. (1989) Nature 337: 376-80
Wright and Voytas (1998) Genetics 149: 703-15

S. RuBisCo small

10

15

20

30

Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCo) catalyzes the initial step in the C3 photosynthetic carbon reduction cycle, adding carbon dioxide to D-ribulose 1,5-bisphosphate to form two molecules of 3-phospho-D-glycerate. RuBisCo is comprised of two subunits, one large which is synthesized in the chloroplast, and one small which is synthesized in the cytoplasm and then transported in to the chloroplast. The expression of the small subunit of RuBisCo is light regulated. Manipulation of these proteins could increase the efficiency of photosynthesis or allow alterations in developmental timing.

Ref: Giuliano et al. (1988) PNAS USA 85: 7089-93 Dedonder et al. (1993) Plant Physiol 101: 801-8

T. Sialyltransf

Members of the CMP-N-acetylneuraminate- β -galactosamide- α -2,3-sialyltransferase family catalyze the following reaction:

CMP-N-acetylneuraminate + β -D-galactosyl-1,3-N-acetyl- α -D-galactosaminyl-R = CMP + α -N-acetylneraminyl-2,3- β -D-galactosyl-1,3-N-acetyl-alpha-D-galactosaminyl-R. These proteins are though to be responsible for the synthesis of the sequence neurac- α -2,3-gal- β -1,3-galnac- found on sugar chains)-linked to threonine or serine and also as a terminal sequence on certain gangliosides in mammalian cells. In plants, glycosyltransferases in the Golgi apparatus synthesize cell wall polysaccharides and elaborate the complex glycans of glycoproteins. Engineering of plant sialyltransferases allows targeting of proteins to particular cellular locations or enables the making of changes in cell wall structure.

Ref: Wee et al. (1998) Plant Cell 10: 1759-68
 Lee et al. (1994) J Biol Chem 269: 10028-33
 Kitagawa and Paulson (1994) J Biol Chem 269: 1394-401

U. Signal

Many plant proteins in this family contain sequences similar to those found in both components of the prokaryotic family of signal transducers known as the two-component systems. This suggests that activation may require a transfer of a phosphate group between

10

15

20

25

30

the transmitter domain and the receiver domain. One family member in Arabidopsis appears to be involved in ethylene (a plant hormone) signal transduction. Other proteins in this family appear to be involved in the regulation of gene transcription under conditions of environmental stress. Signal proteins can be exploited to affect plant growth and development and/or control plant responses to stress conditions such as cold, nutrient availability, etc.

Ref: Chang et al. (1993) Science 262: 539-44
Nagaya et al. (1993) Gene 131: 119-124
Gottfert et al. (1990) PNAS USA 87: 2680-4

V. vMSA

vMSA proteins are major surface antigens presenting on the envelope of various retroviruses. Surface antigens of retroviruses are often involved in tropism of the virus. Plants contain retrovirus-like viruses such as pararetroviruses and retrotransposons (i.e. transposons having long terminal repeats). Plant retrotransposons in particular have been used to create mutants at various loci, thereby permitting gene isolation, gene tagging and the like. Manipulation of plant vMSA proteins enables control of tropism of plant retroviruses that might be used for genetic engineering tools, thus enabling targeting of the virus to particular species and/or tissues of plants.

Ref: Okamoto et al. (1988) J Gen Virol 69: 2575-83 Grandbastien et al. (1989) Nature 337: 376-80 Wright and Voytas (1998) Genetics 149: 703-15

W. zf-CCCH

This family of proteins is defined by having two CX(8)CX(5)CX(3)H-type zinc finger domains. These proteins cover a broad range of functions. For example, the COP1 protein acts as a repressor of photomorphogenesis in darkness; light stimuli abolish this suppressive action. In addition, COP1 protein can function as a negative transcriptional regulator capable of direct interaction with components of the G-protein signaling pathway. As a second example, a zf-CCCH protein identified in Arabidopsis appears to be involved in the resistance to DNA damage induced by UV light and chemical DNA-damaging agents.

Overexpression of this class of proteins permits production of plants that are better suited to adverse environments. Manipulation of expression of zf-CCCH proteins functioning as transcriptional regulators, such as COP1, enables manipulation of some signal transduction pathways.

5

Ref: Pang et al. (1993) Nucleic Acids Res 21: 1647-53 Deng et al. (1992) Cell 71: 791-801

X. zf-RanBP

10

15

Proteins falling within this category contain many X-X-F-G and X-F-X-F-G repeats, and may contain RANBP1-like or PPIase domains. Plant proteins having domains similar to these include PAS1 and GMSTI. PAS1 has been shown to have dramatic developmental affects that appear to be correlated with both cell division and cell wall elongation. GMSTI has high identity to the yeast STI stress-inducible gene and has been shown to be heat inducible. Proteins such as these may be useful for controlling growth and form of development.

20

Ref: Vittorioso et al. (1998) Mol Cell Biol 18: 3034-43 Hernandez Torres et al. (1995) 27: 1221-6

Y. Peptidase M48.

25

30

Proteins belonging to this peptidase family are metalloproteases that bind zinc as a cofactor and are located in the membranes of the endoplasmic reticulum. They function in NH_2 -terminal proteolytic processing, as shown for the yeast STE24 gene product. This gene is required for the correct processing of α -factor, a yeast pheromone. Family M48 peptidases also appear to be required for some prenylation reactions, mediating COOH-terminal CAAX processing. Prenylation reactions are believed to be involved in the regulation of protein-protein and protein-membrane interactions. As an example, RAS GTPase activity is regulated in part by localization to the inner side of the plasma membrane upon prenylation. In plants, proteins from this family could be involved in pollen-stigma interactions such as those mediating self-pollenation vs. outcrossing, or could be members of several secondary metabolism pathways.

10

15

20

25

30

Ref: Fujimura-Kamada et al. (1997) J Cell Biol. 136: 271-85. Tam et al. (1998) J Cell Biol. 142: 635-49.

Z. DNA Pol Viral N

The DNA pol Viral N domain is located at the N-terminal region of DNA polymerase isolated from several retroid viruses such as the Cauliflower Mosaic Virus. The domain motif has also been found in numerous other species from humans to cyanobacteria. In these organisms, this motif seems to be associated with two types of sequences; retrotransposons and mitochondrial genes. In the mitochondrial sequences this domain is potentially involved in the self-splicing conducted by group II introns. Various manipulations of this gene in plants allows control of the numerous retrotransposons endogenous to plant genomes or allows engineering of mitochondrial function, especially to increase efficiency of energy utilization by cells.

REF: Chapdelaine and Bonen (1991) Cell 65: 465-72

Ferat and Miche (1993) Nature 364: 358-61

Wilson et al. (1994) 368: 32-8

Cambareri et al. (1994) 242: 658-65

Gaardner et al. (1981) NAR 9: 2871-2888

Cummings et al. (1990) Curr Genet 17: 375-402

Hattori et al. (1986) Nature 321: 625-8

Aa. Calpain inhib

This domain is found in calpastatin, an inhibitor protein specific for calpain. Calpain is a non-lysosomal calcium-dependent intracellular protease that appears to be involved in the dynamic changes of the cytoskeleton, especially actin-related structures, during early *Drosophila* embryogenesis [1]. Calpastatins co-exist in cells with calpains and the subcellular distribution of calpastatin is thought to be important to calpain regulation [2]. In plants calpains and calpastatins could be involved in embryogenesis and non-embryogenic organ reiteration. Mutations occurring in calpain inhibitor repeat domains would produce developmental abnormalities such as abnormal leaf, root or flower development.

Refs

25

30

- 1 Emori Y and Saigo K (1994) J Biol Chem 269: 25137-42.
- Mellgren RL, Lane RD, Mericle MT (1989) Biochim Biophys Acta 999: 71-77.

Ab. chorismate bind

Chorismate binding domains are present in plant anthranilate synthase (AS) genes. AS genes catalyze the first step in the biosynthesis of tryptophan by converting chorismate and L-glutamine to anthranilate, pyruvate and L-glutamate. Some of these genes are involved in feedback inhibition by tryptophan [1] while some are feedback insensitive [2]. In Arabidopsis, two AS genes have overlapping, but different distributions. One of these AS genes is induced by wounding and bacterial pathogen infiltration [1]. Mutations in the chorismate binding domain would affect the production of tryptophan and could influence the plant's defense system. AS gene products can be used for *in vitro* synthesis of tryptophan and tryptophan derivatives.

15 Refs

5

10

- 1 Niyogi KK, Fink GR (1992) Plant Cell 4: 721-33.
- 2 Song HS, Brotherton JE, Gonzales RA, Wilholm JM (1998) Plant Physiol 117:533-43.

Ac. late protein L2

Papillomaviruses are encapsulated double stranded DNA viruses. Plants are susceptible to infection by double stranded DNA viruses such as Cauliflower Mosaic virus (CaMV). The coat proteins in these plant viruses are critical to the virus life cycle within the plant. For example, the coat protein of CaMV is thought to be involved in intra- and inter-cellular movement within the plant [1]. Engineering of proteins having similarity to papillomavirus coat proteins may enable the production of plants having better resistance to natural plant double stranded DNA viruses.

Refs

1 Thompson SR, Melcher U (1993) J Gen Virol 74: 1141-8.

Ad. Peptidase M41

Proteins belonging to this peptidase family are metalloproteases that bind zinc as a cofactor and are integral membrane proteins. They seem to be involved in the degradation of carboxy-

terminal-tagged cytoplasmic proteins. In plants, these proteins are located in the thylakoid membranes of the chloroplasts, their expression is light regulated and they are thought to be involved in degradation of soluble stromal proteins and turn-over of thylkoid proteins [1]. Manipulation of expression and structure of these proteins would have effects on the efficiency of photosynthesis and the development of chloroplasts.

Refs

5

10

15

25

30

Lindahl M, Tabak s, Cseke L, Pichersky E, Andersson B, Adam Z (1996) J Biol Chem 271: 29329-34.

Ae. UPF0051

There is some evidence that, in plants, proteins in this family are involved in ATP synthesis in chloroplasts [1, 2]. Mutations in these proteins or altering their expression would affect the efficiency of photosynthesis and energy production.

Refs

- 1 Kostrzewa M, Zetsche K (1992) J Mol Biol 227: 961-70.
- 2 Kostrzewa M, Zetsche K (1993) Plant Mol Biol 23: 67-76

20 <u>Af. E7</u>

Papillomaviruses are encapsulated double stranded DNA viruses. The Papillomavirus early protein 7 (E7) is known as a potent immortalizing and transforming agent. Transformation by E7 is thought to be mediated by the physical association of E7 with cellular proteins regulating entry into the cell cycle [1]. The result is entry into the cell cycle and suppression of terminal differentiation in mammalian cells. Thus, engineering of proteins having similarity to papillomavirus E7 protein enables the production of plants having altered cellular proliferation characteristics and possibly altered morphology. For example, overexpression of E7-like proteins would be expected to result in proliferation of cells of the tissue in which the E7 protein is expressed, perhaps with suppression of differentiation events. Thus, for example, overexpression of E7-like proteins in meristem cells can result in taller plants and suppression of leafing and/or flowering.

Refs

1 Zwerschke W, Jansen-Durr P Adv Cancer Res 2000;78:1-29

Ag. Peptidase U7

This protein is known to be an integral membrane protein in the cyanobacterium Synechocystis where it functions to digest cleaved signal peptides [1]. This activity is necessary to maintain proper secretion of mature proteins across the membrane. In higher plants this protein may be present in the plastid or chloroplast membranes where it would function by enabling protein movement into and out of the chloroplasts. Mutations in this protein would be expected to affect the development of plastids, including chloroplasts, or alter the energy transfer system within the chloroplasts, thereby affecting growth and development.

Refs

5

10

15

20

25

30

1 Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S (1996) DNA Res 3:109-36.

Ah. 5'-3' Exonuclease

The 5'-3' exonuclease domain is one found in bacterial DNA polymerases I and in yeast DNA repair enzymes such as Exonuclease I. Yeast Exo I is involved in mitotic recombination and also includes a domain that interacts with the mismatch repair protein MSH2. The 5'-3' exonuclease domain is also present in XPG DNA repair enzymes in humans and in yeast RAD9 protein. Defects in XPG proteins result in Xeroderma Pigmentosum. Thus defects in 5'-3' exonuclease domain-containing proteins in plants are expected to lead to defects in DNA repair and corresponding high spontaneous and inducible mutation rates. Consensus sequence:

IMKKKLLLVDGSSLAFRAFFALPPLTNSAGEPTNAVYGFLKMLIKLIEQEQPTHIAVV FDAKAKTFRHELYEGYKAGRAP

TPDELREQIPLIKELLDALGIPLLEVAGYEADDVIGTLAKLAEKEGYEVLIVTGDRDLL QLVSDHVTVIITKKGIAEFTL

 ${\tt FTPEAVIEKYGLTPEQIIDYKALMGDSSDNIPGVKGIGEKTAAKLLQEYGSLEGIYANL}\\ {\tt DKLKGKKLREKLLAHKEDAKL}$

SRDLATIKTDVPLDLTLDDLRLPDPDRDALDLLFDE



Fiorentini P. et al. RT. Mol. Cell. Biol. 17:2764-2773(1997). Tishkoff et al. Cancer Res. 0:0-0(1998). Macinnes M.A. et al. Mol. Cell. Biol. 13:6393-6402(1993).



Table A			
		Full Name	Description
3_5_exonuclease		3'-5' exonuclease	Accession number: PF01612 Definition: 3'-5' exonuclease
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
		!	Source of seed members: Pfam-B_659 (release 4.1)
			Gathering cutoffs: -11 -11
			Trusted cutoffs: -10.70 -10.70
			Noise cutoffs: -24.50 -24.50 HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
	1		Reference Number: [1]
			Reference Medline: 85137890
			Reference Title: Structure of large fragment of Escherichia
			coli DNA
			Reference Title: polymerase I complexed with dTMP. Reference Author: Ollis DL, Brick P, Hamlin R, Xuong NG,
			Steitz TA;
			Reference Location: Nature 1985;313:762-766.
			Reference Number: [2]
			Reference Medline: 98060913
			Reference Title: The proofreading domain of Escherichia
			coli DNA polymerase Reference Title: I and other DNA and/or RNA exonuclease
			domains.
			Reference Author: Moser MJ, Holley WR, Chatterjee A, Mian
			IS;
			Reference Location: Nucleic Acids Res 1997;25:5110-5118. Reference Number: [3]
			Reference Medline: 98361165
			Reference Title: Replication focus-forming activity 1 and the
			Werner
			Reference Title: syndrome gene product
			Reference Author: Yan H, Chen CY, Kobayashi R, Newport
			J; Reference Location: Nat Genet 1998;19:375-378.
			Reference Number: [4]
			Reference Medline: 97434221
			Reference Title: The Werner syndrome protein is a DNA
			helicase. Reference Author: Gray MD, Shen JC, Kamath-Loeb AS,
			Reference Author: Gray MD, Shen JC, Kamath-Loeb AS, Blank A, Sopher BL,
			Reference Author: Martin GM, Oshima J, Loeb LA;
			Reference Location: Nat Genet 1997;17:100-103.
			Reference Number: [5]
			Reference Medline: 97370026
			Reference Title: DNA helicase activity in Werner's syndrome gene product
			Reference Title: synthesized in a baculovirus system.
	1		Reference Author: Suzuki N, Shimamoto A, Imamura O,
			Kuromitsu J, Kitao S,
	1		Reference Author: Goto M, Furuichi Y; Reference Location: Nucleic Acids Res 1997;25:2973-2978.
			Reference Location: Nucleic Acids Res 1997;25:2973-2978. Database Reference: SCOP; 1dpi; fa; [SCOP-USA][CATH-
			PDBSUM)
			Database Reference INTERPRO; IPR002562;
			Database Reference PDB; 1kfd; 348; 518;
			Database Reference PDB; 1d8y A; 348; 518;
			Database Reference PDB; 1d9d A; 348; 518; PDB; 1d9f A; 348; 518;
			Database Reference PDB; 1kfs A; 348; 518;
			Database Reference PDB; 1kln A; 348; 518;
			Database Reference PDB; 1krp A; 348; 518;
			Database Reference PDB; 1ksp A; 348; 518;
			Database Reference PDB; 1qsl A; 348; 518; PDB; 2kfn A; 348; 518;
			Database Reference PDB; 2kfz A; 348; 518;
			Database Reference PDB; 2kzm A; 348; 518;
			Database Reference PDB; 2kzz A; 348; 518;
			Comment: This domain is responsible for the 3'-5'
			exonuclease proofreading Comment: activity of E. coli DNA polymerase I (poll)
			Comment: activity of E. coli DNA polymerase I (poli) and other enzymes,
	L		January Completion





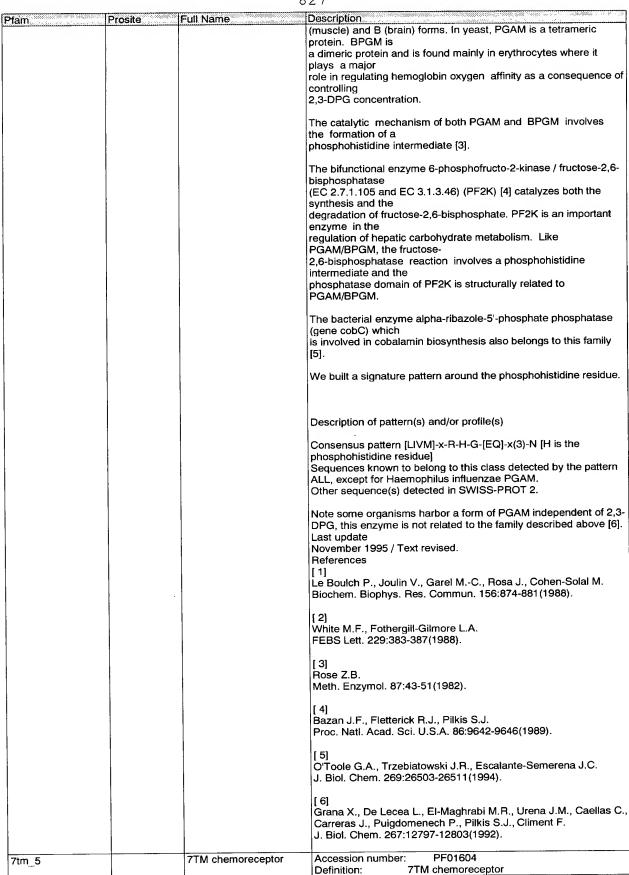
Pfam	Prosite	Full Name	Description
ICHITE	1100110		Comment: it catalyses the hydrolysis of unpaired or
			mismatched nucleotides.
			Comment: This domain consists of the amino-terminal
			half of the Klenow fragment
			Comment: in E. coli poll it is also found in the Werner
			syndrome helicase
			Comment: (WRN), focus forming activity 1 protein
			(FFA-1) and ribonuclease D Comment: (RNase D).
			Y
			Comment: Werner syndrome is a human genetic disorder causing premature aging;
	!		
	:		Comment: the WRN protein has helicase activity in the 3'-5' direction [4,5].
			Comment: The FFA-1 protein is required for formation
			of a replication foci
			Comment: and also has helicase activity; it is a
			homologue of the WRN
			Comment: protein [3].
			Comment: RNase D is a 3'-5' exonuclease involved in
			tRNA processing.
			Comment: Also found in this family is the autoantigen
			PM/Scl thought to be
			Comment: involved in polymyositis-scleroderma
		1	overlap syndrome.
			Number of members: 41
			TAGINGO OF MONIBORS.
3HCDH	PDOC00065	3-hydroxyacyl-CoA	3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) (HCDH) [1] is
SHODH	F D C C C C C C C C C C C C C C C C C C	dehydrogenase signature	an enzyme involved
		deriyaregenase orginarare	in fatty acid metabolism, it catalyzes the reduction of 3-
			hydroxyacyl-CoA to
			3-oxoacyl-CoA. Most eukaryotic cells have 2 fatty-acid beta-
		ľ	oxidation systems,
			one located in mitochondria and the other in peroxisomes. In
			peroxisomes
			3-hydroxyacyl-CoA dehydrogenase forms, with enoyl-CoA
			hydratase (ECH) and
			3,2-trans-enoyl-CoA isomerase (ECI) a multifunctional enzyme
			where the N-
			terminal domain bears the hydratase/isomerase activities and
			the C-terminal
			domain the dehydrogenase activity. There are two mitochondrial
			enzymes: one
			which is monofunctional and the other which is, like its
			peroxisomal
			counterpart, multifunctional.
			In Escherichia coli (gene fadB) and Pseudomonas fragi (gene
			faoA) HCDH is part
			of a multifunctional enzyme which also contains an ECH/ECI
			domain as well as a
			3-hydroxybutyryl-CoA epimerase domain [2].
			A
			The other proteins structurally related to HCDH are:
		1	(50444457)
			- Bacterial 3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157)
	1		which reduces
			3-hydroxybutanoyl-CoA to acetoacetyl-CoA [3].
			- Eye lens protein lambda-crystallin [4], which is specific to
			lagomorphes
			(such as rabbit).
1			
			There are two major region of similarities in the sequences of
			proteins of the
			HCDH family, the first one located in the N-terminal, corresponds
1			to the NAD-
			binding site, the second one is located in the center of the
			sequence. We have
			chosen to derive a signature pattern from this central region.
	1		
			Description of pattern(s) and/or profile(s)
I		1	·
			Consensus pattern [DNE]-x(2)-[GA]-F-[LIVMFY]-x-[NT]-R-x(3)-



		O	<u> </u>
Pfam	Prosite		Description [PA]-[LIVMFY](2)- x(5)-[LIVMFYCT]-[LIVMFY]-x(2)-[GV] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update July 1998 / Pattern and text revised. References [1] Birktoff J.J., Holden H.M., Hamlin R., Xuong NH., Banaszak L.J. Proc. Natl. Acad. Sci. U.S.A. 84:8262-8266(1987).
			[2] Nakahigashi K., Inokuchi H. Nucleic Acids Res. 18:4937-4937(1990).
			[3] Mullany P., Clayton C.L., Pallen M.J., Slone R., Al-Saleh A., Tabaqchali S. FEMS Microbiol. Lett. 124:61-67(1994).
			[4] Mulders J.W.M., Hendriks W., Blankesteijn W.M., Bloemendal H., de Jong W.W. J. Biol. Chem. 263:15462-15466(1988).
4HPPD_C		4-hydroxyphenylpyruvate dioxygenase C terminal domain	Accession number: PF01626 Definition: 4-hydroxyphenylpyruvate dioxygenase C terminal domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1116 (release 4.1) Gathering cutoffs: -35 -35 Trusted cutoffs: -25.80 -25.80 Noise cutoffs: -44.90 -44.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93279307 Reference Title: Human 4-hydroxyphenylpyruvate dioxygenase. Primary Reference Title: structure and chromosomal localization of the gene. Reference Author: Ruetschi U, Dellsen A, Sahlin P, Stenman G, Rymo L, Reference Author: Ruetschi U, Dellsen A, Sahlin P, Stenman G, Rymo L, Reference Location: INTERPRO; IPR002887; 4-Hydroxyphenylpyruvic acid dioxygenase (HPD) is an important enzyme Comment: 4-Hydroxyphenylpyruvic acid dioxygenase (HPD) is an important enzyme Comment: this enzyme in humans and mice leads to hereditary tyrosinemia type 3. Comment: this enzyme in humans and mice leads to hereditary tyrosinemia type 3. Comment: molecule a candidate for a functional role in the catalytic process Comment: [1]. This region is found as a separate protein Swiss:Q49717 that Comment: is somewhat different from HPD and may have a different but related Comment: genetic deficiency in protein function (Unpublished observation Bateman A). Number of members: 28



Pfam	Prosite	Full Name	Description
5_3_exonuclease		5'-3' exonuclease domain	The 5'-3' exonuclease domain is one found in bacterial DNA polymerases I and in yeast DNA repair enzymes such as Exonuclease I. Yeast Exo I is involved in mitotic recombination and also includes a domain that interacts with the mismatch repair protein MSH2. The 5'-3' exonuclease domain is also present in XPG DNA repair enzymes in humans and in yeast RAD9 protein. Defects in XPG proteins result in Xeroderma Pigmentosum. Thus defects in 5'-3' exonuclease domain-containing proteins in plants are expected to lead to defects in DNA repair and corresponding high spontaneous and inducible mutation rates. Consensus sequence: IMKKKLLLVDGSSLAFRAFFALPPLTNSAGEPTNAVYGFLKMLIK LIEQEQPTHIAVVFDAKAKTFRHELYEGYKAGRAP TPDELREQIPLIKELLDALGIPLLEVAGYEADDVIGTLAKLAEKEG YEVLIVTGDRDLLQLVSDHVTVIITKKGIAEFTL FTPEAVIEKYGLTPEQIIDYKALMGDSSDNIPGVKGIGEKTAAKLL QEYGSLEGIYANLDKLKGKKLREKLLAHKEDAKL SRDLATIKTDVPLDLTLDDLRLPDPDRDALDLLFDE
			Fiorentini P. et al. RT. Mol. Cell. Biol. 17:2764-2773(1997). Tishkoff et al. Cancer Res. 0:0-0(1998). Macinnes M.A. et al. Mol. Cell. Biol. 13:6393-6402(1993).
60s_ribosomal		60s Acidic ribosomal protein	Accession number: PF00428 Definition: 60s Acidic ribosomal protein Author: Finn RD Alignment method of seed: Clustalw Source of seed members: Pfam-B_151 (release 1.0) Gathering cutoffs: 17.80 Trusted cutoffs: 17.80 17.80 Noise cutoffs: 9.30 9.30 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 96282699 Reference Title: ribosome stalk. New structural and functional aspects. Reference Author: Remacha M, Jimenez-Diaz A, Santos C, Briones E, Zambrano R, Reference Author: Ballesta JP; Reference Location: Database Reference Database reference: Database reference: Comment: PFAMB; PB002218; This family includes archaebacterial L12, eukaryotic P0, P1 and P2. Number of members: 109
6PF2K	PDOC00158	Phosphoglycerate mutase family phosphohistidine signature	Phosphoglycerate mutase (EC 5.4.2.1) (PGAM) and bisphosphoglycerate mutase (EC 5.4.2.4) (BPGM) are structurally related enzymes which catalyze reactions involving the transfer of phospho groups between the three carbon atoms of phosphoglycerate [1,2]. Both enzymes can catalyze three different reactions, although in different proportions: - The isomerization of 2-phosphoglycerate (2-PGA) to 3-phosphoglycerate (3-PGA) with 2,3-diphosphoglycerate (2,3-DPG) as the primer of the reaction. - The synthesis of 2,3-DPG from 1,3-DPG with 3-PGA as a primer. - The degradation of 2,3-DPG to 3-PGA (phosphatase EC 3.1.3.13 activity). In mammals, PGAM is a dimeric protein. There are two isoforms of PGAM: the M





			328
Pfam	Prosite	Full Name	Description
			Author: Bateman A
		İ	Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_942 (release 4.1) Gathering cutoffs: -46 -46
			Trusted cutoffs: -44.30 -44.30
			Noise cutoffs: -47.80 -47.80
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 98248686
			Reference Title: Two large families of chemoreceptor genes
		1	in the nematodes
			Reference Title: Caenorhabditis elegans and Caenorhabditis
			briggsae reveal
			Reference Title: extensive gene duplication, diversification,
			movement, and
			Reference Title: intron loss. Reference Author: Robertson HM;
l			Reference Location: Genome Res 1998;8:449-463.
			Database Reference INTERPRO; IPR003003;
			Comment: This large family of proteins are related to
			7tm 1.
			Comment: They are 7 transmembrane receptors. This
			family does not
			Comment: include all known members, as there are
			problems with
			Comment: overlapping specificity with 7tm_1.
			Comment: This family is greatly expanded in the
			nematode worm C.
			Comment: elegans.
			Number of members: 180
As trops		Transmembrane amino	Accession number: PF01490
Aa_trans		acid transporter protein	Definition: Transmembrane amino acid transporter
		acid transporter protein	protein
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B 419 (release 4.0)
			Gathering cutoffs: 25 25
	1		Trusted cutoffs: 150.80 150.80
			Noise cutoffs: 3.60 3.60
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
		i	Reference Medline: 98007977 Reference Title: Identification and characterization of the
	1		Reference Title: Identification and characterization of the vesicular GABA
			Reference Title: transporter.
			Reference Author: McIntire SL, Reimer RJ, Schuske K,
	1		Edwards RH, Jorgensen
	1		Reference Author: EM;
	1		Reference Location: Nature 1997;389:870-876.
			Database Reference INTERPRO; IPR002422;
			Database reference: PFAMB; PB020912;
			Comment: This transmembrane region is found in
1			many amino acid transporters
			Comment: including UNC-47 and MTR. UNC-47
			encodes a vesicular amino butyric acid
			Comment: (GABA) transporter, (VGAT). UNC-47 is
			predicted to have 10 transmembrane
			Comment: domains Swiss:P34579 [1]. MTR is a N system amino acid transporter system
			Comment: protein involved in methyltryptophan
			resistance Swiss:P38680.
			Comment: Other members of this family include proline
		1	transporters and amino
			Comment: acid permeases.
			Number of members: 50
ABC_tran	PDOC00185	ABC transporters family	On the basis of sequence similarities a family of related
		signature	ATP-binding
1			proteins has been characterized [1 to 5]. These proteins are
			associated with a
			variety of distinct biological processes in both prokaryotes and

		8	329
Pfam	Prosite	Full Name	Description
			eukaryotes,
			but a majority of them are involved in active transport of small hydrophilic
1			molecules across the cytoplasmic membrane. All these
			proteins share a
			conserved domain of some two hundred amino acid residues,
			which includes an ATP-binding site. These proteins are collectively known as ABC
			transporters.
			Proteins known to belong to this family are listed below
			(references are only
			provided for recently determined sequences).
			In prokaryotes:
			- Active transport systems components: alkylphosphonate uptake(phnC/phnK/
			phnL); arabinose (araG); arginine (artP); dipeptide (dciAD;dppD/dppF);
			ferric enterobactin (fepC); ferrichrome (fhuC); galactoside
			(mglA); glutamine (glnQ); glycerol-3-phosphate (ugpC); glycine
			betaine/L-proline (proV); glutamate/aspatate (gltL); histidine (hisP); iron(III)
			(sfuC),
			iron(III) dicitrate (fecE); lactose (lacK);
	-		leucine/isoleucine/valine (braF/braG;livF/livG); maltose (malK); molybdenum (modC);
ļ			nickel (nikD/
			nikE); oligopeptide (amiE/amiF;oppD/oppF); peptide
			(sapD/sapF); phosphate
			(pstB); putrescine (potG); ribose (rbsA); spermidine/putrescine (potA);
			sulfate (cysA); vitamin B12 (btuD).
			Handhair Vardada in ann ad an daire bh Dann Barri Na B
			Hemolysin/leukotoxin export proteins hlyB, cyaB and lktB. Colicin V export protein cvaB.
			- Lactococcin export protein lcnC [6].
			- Lantibiotic transport proteins nisT (nisin) and spaT (subtilin).
			- Extracellular proteases B and C export protein prtD.
			- Alkaline protease secretion protein aprD. - Beta-(1,2)-glucan export proteins chvA and ndvA.
			- Haemophilus influenzae capsule-polysaccharide export protein
			bexA.
			- Cytochrome c biogenesis proteins ccmA (also known as cycV
			and helA) Polysialic acid transport protein kpsT.
			- Cell division associated ftsE protein (function unknown).
			- Copper processing protein nosF from Pseudomonas stutzeri.
			- Nodulation protein nodI from Rhizobium (function unknown).
			- Escherichia coli proteins cydC and cydD. - Subunit A of the ABC excision nuclease (gene uvrA).
			- Erythromycin resistance protein from Staphylococcus
			epidermidis (gene
			msrA). Tylogin registance protein from Streetenwees fradice (gene thrC)
			- Tylosin resistance protein from Streptomyces fradiae (gene tlrC) [7].
			- Heterocyst differentiation protein (gene hetA) from Anabaena PCC 7120.
			- Protein P29 from Mycoplasma hyorhinis, a probable
			component of a high
			affinity transport system.
			- yhbG, a putative protein whose gene is linked with ntrA in many bacteria
			such as Escherichia coli, Klebsiella pneumoniae,
			Pseudomonas putida,
			Rhizobium meliloti and Thiobacillus ferrooxidans.
			- Escherichia coli and related bacteria hypothetical proteins yabJ, yadG,
			yabb, yadG, yagC, ybbA, ycjW, yddA, yehX, yejF, yheS, yhiG, yhiH, yjcW,
			yijK, yojl,
			yrbF and ytfR.
			In aukanyotas:
		<u> </u>	In eukaryotes:



		8	30
Pfam I	Prosite	Full Name	Description
			- The multidrug transporters (Mdr) (P-glycoprotein), a family of closely
			related proteins which extrude a wide variety of drugs out of the cell (for
			a review see [8]) Cystic fibrosis transmembrane conductance regulator (CFTR), which is most
			probably involved in the transport of chloride ions Antigen peptide transporters 1 (TAP1, PSF1, RING4, HAM-
			1, mtp1) and 2 (TAP2, PSF2, RING11, HAM-2, mtp2), which are involved in the transport of
			antigens from the cytoplasm to a membrane-bound compartment for
			association with MHC class I molecules 70 Kd peroxisomal membrane protein (PMP70) ALDP, a peroxisomal protein involved in X-linked
			adrenoleukodystrophy [9] Sulfonylurea receptor [10], a putative subunit of the B-cell ATP- sensitive
			potassium channel Drosophila proteins white (w) and brown (bw), which are
			involved in the import of ommatidium screening pigments Fungal elongation factor 3 (EF-3).
			- Yeast STE6 which is responsible for the export of the a-factor pheromone.
			 Yeast mitochondrial transporter ATM1. Yeast MDL1 and MDL2. Yeast SNQ2.
			- Yeast sporidesmin resistance protein (gene PDR5 or STS1 or YDR1).
			 Fission yeast heavy metal tolerance protein hmt1. This protein is probably involved in the transport of metal-bound phytochelatins.
			- Fission yeast brefeldin A resistance protein (gene bfr1 or hba2) Fission yeast leptomycin B resistance protein (gene pmd1) mbpX, a hypothetical chloroplast protein from Liverwort Prestalk-specific protein tagB from slime mold. This protein consists of
			two domains: a N-terminal subtilase catalytic domain (see < <u>PDOC00125</u> >) and a C-terminal ABC transporter domain.
			As a signature pattern for this class of proteins, we use a conserved region
			which is located between the 'A' and the 'B' motifs of the ATP- binding site.
			Consensus pattern [LIVMFYC]-[SA]-[SAPGLVFYKQH]-G-[DENQMW]- [KRQASPCLIMFW]- [KRNQSTAVM]-[KRACLVM]-[LIVMFYPAN]- {PHY}-[LIVMFW]- [SAGCLIVP]-{FYWHP}-{KRHP}- [LIVMFYWSTA] Sequences known to belong to this class detected by the pattern ALL, except for 25 sequences. Other
			sequence(s) detected in SWISS-PROT 42. Note the ATP-binding region is duplicated in araG, mdl, msrA, rbsA, tlrC, uvrA, yejF, Mdr's, CFTR, pmd1 and in EF-3. In some of those proteins, the above pattern only detect one of the two copies of the domain. Note the proteins belonging to this family also contain one or two copies of the ATP-binding motifs 'A' and 'B' (see < PDC00017>).
			July 1998 / Text revised. [1] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P.
			J. Bioenerg. Biomembr. 22:571-592(1990). [2] Higgins C.F., Gallagher M.P., Mimmack M.M., Pearce S.R.
			BioEssays 8:111-116(1988). [3] Higgins C.F., Hiles I.D., Salmond G.P.C., Gill D.R., Downie J.A.,
			Evans I.J., Holland I.B., Gray L., Buckels S.D., Bell A.W., Hermodson M.A. Nature 323:448-450(1986).

			331
Pfam	Prosite	Full Name	Description
			[4] Doolittle R.F., Johnson M.S., Husain I., van Houten B., Thomas D.C., Sancar A.
,			Nature 323:451-453(1986). [5] Blight M.A., Holland I.B.
			Mol. Microbiol. 4:873-880(1990).
			Stoddard G.W., Petzel J.P., van Belkum M.J., Kok J., McKay L.L. Appl. Environ. Microbiol. 58:1952-1961(1992).
			Rosteck P.R. Jr., Reynolds P.A., Hershberger C.L. Gene 102:27-32(1991).
			Gottesman M.M., Pastan I. J. Biol. Chem. 263:12163-12166(1988).
			Valle D., Gaertner J. Nature 361:682-683(1993). [10]
			Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV, Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguy K., Bryan J., Nelson D.A. Science 268:423-426(1995).
ABC2_membrane	PDOC00692	ABC-2 type transport system integral membrane proteins signature	Integral membrane components of a number of bacterial active transport systems have been shown to be evolutionary related and to form a distinct family
			[1,2]. These proteins are:
			Escherichia coli kpsM, involved in polysialic acid export. Haemophilus influenzae bexB, involved in polyribosylribitol phosphate
			capsule polysaccharide export.
			- Salmonella typhi vexB, involved in translocation of the Vi polysaccharide.
			- Neisseria meningitidis ctrC, involved in polyneuraminic acid capsule
			polysaccharide export Rhizobiacae nodulation protein J (gene nodJ), probably
			involved in exporting a modified beta-1,4-linked N-acetylglucosamine
			oligosaccharide Streptomyces peucetius drrB, involved in exporting the
			antibiotics daunorubicin and doxorubicin.
			- Klebsiella pneumoniae O-antigen exprt system protein rfbA Yersinia enterocolitica O-antigen exprt system protein rfbD Escherichia coli hypothetical protein yadH.
			- Escherichia coli hypothetical protein yhhJ.
			The molecular size of these proteins is around 30 Kd. They are thought to
			contain six transmembrane regions. They either form homooligomeric channels or
			associate with another type of transmembrane protein to form heteroligomers.
			Transport systems in which they participate are energized by an ATP-binding
			protein that belongs to the ABC transporter family. The designation 'ABC-2'
			has been proposed [1] for these transport systems.
			As a signature pattern, we selected a conserved region located in the C-terminal section of these proteins.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIMST]-x(2)-[LIMW]-x(2)-[LIMCA]-[GSTC]-x-[GSAIV]-x(6)- [LIMGA]-[PGSNQ]-x(9,12)-P-[LIMFT]-x-[HRSY]-

			332
Pfam	Prosite	Full Name	Description
			x(5)-[RQ] Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT 2. Last update
			November 1997 / Pattern and text revised. References
			[1] Reizer J., Reizer A., Saier M.H. Jr. Protein Sci. 1:1326-1332(1992).
			[2] Vazquez M., Santana O., Quinto C. Mol. Microbiol. 8:369-377(1993).
ABC-3		ABC 3 transport family	Members of this family include receptors that mediate transmembrane signalling. These receptors can bind to a number of factors including: amphiregulin, epidermal growth factor, gp30, heparin-binding egf, insulin, insulin-like growth factor I and II, neuregulins, transforming growth factor-alpha and, and vaccinia virus growth
			Signal transduction is mediated by catalytic activity of tyrosine kinase, such as ATP + A protein tyrosine = ADP + protein tyrosine phosphate. Typically, such signal transduction have been implicated in metabolic and developmental changes, including cell fate and differentiation. Examples include instruction of follicle cells to follow a dorsal pathway of development rather than the default ventral pathway. may also bind the spitz protein. References describing these family members and their biological activities:
			Abbot et al., J. Biol. Chem. 267:10759-10763(1992); Araki et al., J. Biol. Chem. 262:16186-16191(1987); Aroian et al., EMBO J. 13:360-366(1994); Aroian et al., Nature 348:693-699(1990); Barbetti et al., Diabetes 41:408-415(1992); Bargmann et al., Nature 319:226-230(1986); Cama et al., J. Biol. Chem. 268:8060-8069(1993); Cama et al., J. Clin. Endocrinol. Metab. 73:894-901(1991); Carrera et al., Hum. Mol. Genet. 2:1437-1441(1993); Clifford et al., Genetics 137:531-550(1994); Cocozza et al., Diabetes 41:521-526(1992); Cooke et al., Biochem. Biophys. Res. Commun. 177:1113-1120(1991); Coussens et al., Science 230:1132-1139(1985); Dickens et al., Biochem. Biophys. Res. Commun. 186:244-250(1992); Ebina et al., Cell 40:747-758(1985); Ebina et al., Genomics 15:426-429(1993); Elbein et al., Diabetes 42:429-434(1993); Elbein, Diabetes 38:737-743(1989); Fujita-Yamaguchi et al., Protein Seq. Data Anal. 1:3-6(1987); Gullick et al., EMBO J. 11:43-48(1992); Haruta et al., Diabetes 42:1837-1844(1993); Hubbard et al., EMBO J. 16:5572-5581(1997). Hubbard et al., Nature 372:746-754(1994); Iwanishi et al., Diabetologia 36:414-422(1993); Kadowaki et al., J. Clin. Invest. 86:254-264(1990); Kadowaki et al., Science 240:787-790(1988); Kim et al., Diabetologia 35:261-266(1992); Klinkhamer et al., EMBO J. 8:2503-2507(1989); Kusari et al., J. Biol. Chem. 268:51272-11277(1993); Lee et al., Oncogene 8:3403-3410(1993); Lesokhin et al., Dev. Biol. 205:129-144(1999); Livneh et al., Cell 40:599-607(1985). Longo et al., Proc. Natl. Acad. Sci. U.S.A. 90:60-64(1993); McKeon et al., Mol. Endocrinol. 4:647-656(1990); Moller et al., J. Biol. Chem. 265:14979-14985(1990); Moller et al., Science 245:66-68(1989); Raz et al., Genetics 129:191-201(1991). Sakai et al., J. Mol. Biol. 256:548-555(1996); Schaeffer et al., Biochem. Biophys. Res. Commun. 189:650-653(1992); Schejter et al., Cell 46:1091-1101(1986); Seino et al., Biochem. Biophys.
; 			Res. Commun. 159:312-316(1989); Seino et al., Diabetes 39:123-128(1990); Semba et al., Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985); Shier et al., J. Biol. Chem. 264:14605-14608(1989); Taira et al., Science 245:63-66(1989); Tewari et al., J. Biol.



		8	333
Pfam	Prosite	Full Name	Description
			Chem. 264:16238-16245(1989); Ullrich et al., Nature 313:756-
			761(1985).
			Ullrich et al., EMBO J. 5:2503-2512(1986); van der Vorm et al.,
			Diabetologia 36:172-174(1993); van der Vorm et al., J. Biol.
			Chem. 267:66-71(1992); Wadsworth et al., Nature 314:178-
			180(1985); White et al., Cell 54:641-649(1988); Xu et al., J. Biol.
			Chem. 265:18673-18681(1990); Yamamoto et al., Nature
			319:230-234(1986); and Yoshimasa et al., Science 240:784-
			787(1988).
ACAT		Sterol O-acyltransferase	Accession number: PF01800
		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	Definition: Sterol O-acyltransferase
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_1454 (release 4.2)
		İ	Gathering cutoffs: 25 25
			Trusted cutoffs: 112.80 112.80
			Noise cutoffs: -128.10 -128.10
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
		•	Reference Number: [1] Reference Medline: 98434592
			Reference Title: Characterization of two human genes
			encoding acyl coenzyme
			Reference Title: A:cholesterol acyltransferase-related
			enzymes.
			Reference Author: Oelkers P, Behari A, Cromley D,
			Billheimer JT, Sturley SL;
			Reference Location: J Biol Chem 1998;273:26765-26771.
			Reference Number: [2]
			Reference Medline: 98434590
			Reference Title: Identification of a form of acyl-
i			Reference Title: acyltransferase specific to liver and
			intestine in nonhuman
			Reference Title: primates.
			Reference Author: Anderson RA, Joyce C, Davis M, Reagan
			JW, Clark M, Shelness
			Reference Author: GS, Rudel LL;
			Reference Location: J Biol Chem 1998;273:26747-26754.
			Reference Number: [3]
			Reference Medline: 96243137 Reference Title: 96243137 Sterol esterification in yeast: a two-gene
			Reference Title: Sterol esterification in yeast: a two-gene process.
			Reference Author: Yang H, Bard M, Bruner DA, Gleeson A,
			Deckelbaum RJ,
			Reference Author: Aljinovic G, Pohl TM, Rothstein R, Sturley
			SL;
			Reference Location: Science 1996;272:1353-1356.
İ			Database Reference INTERPRO; IPR002688;
			Comment: Sterol O-acyltransferases or acyl-
į			coa:cholesterol acyltransferase
			Comment: (ACAT) EC:2.3.1.26 is a transmembrane
			protein that catalyses the
			Comment: esterification of cholesterol to its cholesterol
			ester storage Comment: form.
			Number of members: 21
ACPS		4'-phosphopantetheinyl	Accession number: PF01648
		transferase superfamily	Definition: 4'-phosphopantetheinyl transferase
			superfamily
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_1679 (release 4.1)
			Gathering cutoffs: 0 0
			Trusted cutoffs: 0.60 0.60 Noise cutoffs: -4.00 -4.00
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 96027548
			Reference Title: Cloning, overproduction, and
			her.



			334	
Pfam	Prosite	Full Name	Description	
	7.20		characterization of the	
		II.	Reference Title: synthase.	Escherichia coli holo-acyl carrier protein
		1	Reference Author:	Lambalot RH, Walsh CT;
		j	Reference Location:	
			Reference Number:	[2]
			Reference Medline: Reference Title:	97144264
			phosphopantetheiny	A new enzyme superfamily - the
			Reference Title:	transferases.
			Reference Author: Zuber P, LaCelle M,	Lambalot RH, Gehring AM, Flugel RS,
			Reference Author:	Marahiel MA, Reid R, Khosla C, Walsh
			Reference Location: Reference Number:	Chem Biol 1996;3:923-936.
	1		Reference Medline:	[3] 10581256
			Reference Title:	Crystal structure of the surfactin
			synthetase-activating	
			Reference Title:	enzyme sfp: a prototype of the 4'-
			phosphopantetheinyl Reference Title:	transferase superfamily [In Process
			Citation]	Devided K. M. C. M. D. M
			Reference Author: R;	Reuter K, Mofid MR, Marahiel MA, Ficner
			Reference Location:	EMBO J 1999;18:6823-6831.
			Database Reference	INTERPRO; IPR002582;
			Database reference: Database reference:	PFAMB; PB007908; PFAMB; PB041384:
			Comment:	Members of this family transfers the
			Comment:	4'-phosphopantetheine (4'-PP) moiety from
			coenzyme A (CoA) to	
			Comment: translational	the invariant serine of pp-binding. This post-
			Comment:	modification renders holo-ACP capable of
			acyl group activation	meaniferrential of files from dapasie of
			Comment: of 4'-PP [1].	via thioesterification of the cysteamine thiol
			Comment:	This superfamily consists of two subtypes:
		i	The ACPS type Comment:	such as Swiss:P24224 and the Sfp type
			such as Swiss:P3913 Comment:	5. The structure of the Sfp type is known [3],
			which shows the	active site accommodates a magnesium ion.
			The most highly	
			involved in binding	conserved regions of the alignment are
			Comment: Number of members:	the magnesium ion. 46
ACT				
]'		Accession number: Definition: A(PF01842 CT domain
			Author: Ba	iteman A
			Alignment method of:	seed: Manual
			Source of seed memb	
			Gathering cutoffs: Trusted cutoffs:	25 0 26.10 0.50
	ļ			24.50 24.50
			HMM build command	line: hmmbuild HMM SEED
			HMM build command	line: hmmcalibrateseed 0 HMM
		- 1	Reference Number: Reference Medline:	[1] 95236205
			Reference Title:	The allosteric ligand site in the Vmax-type
			cooperative Reference Title:	enzyme phoephoglyporete debydaese
			Reference Author:	enzyme phosphoglycerate dehydrogenase. Schuller DJ, Grant GA, Banaszak لنا;
			Reference Location:	Nat Struct Biol 1995;2:69-76.
			Reference Number:	[2]
			Reference Medline: Reference Title:	99241053 Gleaning non-trivial structural, functional
		-	and	- ·
			terative	evolutionary information about proteins by

		835	
Pfam Prosite	Full Name	Description	
		Reference Title:	database searches.
		Reference Author:	Aravind L, Koonin EV;
		Reference Location:	J Mol Biol 1999;287:1023-1040.
		Database Reference	: SCOP; 1psd; fa; [SCOP-USA][CATH-
		Database Reference	INTERPRO; IPR002912;
		Database Reference	,
	j	Database Reference	, . [,,
		Database Reference	
		Database Reference	
		Database reference:	PFAMB; PB001977;
		Database reference:	PFAMB; PB008097;
		Database reference:	PFAMB; PB010480;
		Database reference:	PFAMB; PB011031;
		Database reference: Database reference:	PFAMB; PB031880; PFAMB; PB038464;
		Database reference:	PFAMB; PB040963;
		Database reference:	PFAMB; PB041518;
		Database reference:	PFAMB; PB041667;
		Comment:	This family of domains generally have a
		regulatory role.	3
		Comment:	ACT domains are linked to a wide range of
		metabolic	_
		Comment:	enzymes that are regulated by amino acid
		concentration.	Delegation of ACT described to the control of the c
		Comment:	Pairs of ACT domains bind specifically to a
		particular Comment:	amino acid leading to regulation of the
		linked enzyme.	animo acid leading to regulation of the
		Comment:	The ACT domain is found in:
		Comment:	D-3-phosphoglycerate dehydrogenase
		EC:1.1.1.95 Swiss:P0	
			which is inhibited by serine [1].
			Aspartokinase EC:2.7.2.4 Swiss:P53553,
		which is regulated by	
		Comment: subunit Swiss:P00894	Acetolactate synthase small regulatory
		_	which is inhibited by valine.
		1_	Phenylalanine-4-hydroxylase EC:1.14.16.1
		Swiss:P00439, which	· many man mile i my aroxy table 20.117 1.10.1
		Comment:	is regulated by phenylalanine.
			Prephenate dehydrogenase EC:4.2.1.51
		Swiss:P21203.	
			formyltetrahydrofolate deformylase
		EC:3.5.1.10, Swiss:P: Comment:	,
		inhibited by glycine.	which is activated by methionine and
		_	GTP pyrophosphokinase EC:2.7.6.5
		Swiss:P11585.	GTT Pyrophiosphiokinado Eo.E.7.0.0
		Number of members:	177
Activin_recp	Activin types I and II	Accession number:	PF01064
İ	receptor domain		tivin types I and II receptor domain
		Author: Fin	n RD, Bateman A
ļ			eed: Clustalw_manual ers: Pfam-B 338 (release 3.0)
		Gathering cutoffs:	22 22
		5	23.10 23.10
			1.30 21.20
		HMM build command I	ine: hmmbuild -F HMM SEED
			ine: hmmcalibrateseed 0 HMM
		Reference Number:	[1]
		Reference Medline:	97454714
		Reference Title: pathway.	From receptor to nucleus: the Smad
		Reference Author:	Baker JC, Harland RM;
		Reference Location:	Curr Opin Genet Dev 1997;7:467-473.
		Reference Number:	[2]
		Reference Medline:	94131268
			The TGF-beta superfamily: new members,
		new receptors, and	· · · ·
			new genetic tests of function in different
ŀ		organisms.	
		Reference Author:	Kingsley DM;





			836	
Pfam	Prosite	Full Name	Description	
			Reference Location:	Genes Dev 1994;8:133-146.
			Reference Number: Reference Medline:	[3] 93390967
			Reference Title:	Activin receptor-like kinases: a novel
			subclass of	
			Reference Title:	cell-surface receptors with predicted
			serine/threonine Reference Title:	kinasa aetivity
			Reference Author:	kinase activity. ten Dijke P, Ichijo H, Franzen P, Schulz P,
			Saras J,	terral que en fremje en franzen e
			Reference Author:	Toyoshima H, Heldin CH, Miyazono K;
			Reference Location: Database Reference	Oncogene 1993;8:2879-2887. INTERPRO; IPR000472;
			Database reference:	PFAMB; PB024112;
			Database reference:	PFAMB; PB040755;
			Comment:	This Pfam entry consists of both TGF-beta
	İ		receptor types. Comment:	This is an alignment of the budget life
		İ	cysteine-rich	This is an alignment of the hydrophilic
			Comment:	ligand-binding domains,
			Comment:	Both receptor types, (type I and II) posses a
	İ		9 amino	
			CCX{4-5}CN.	acid cysteine box, with the the consensus
			Comment:	The type I receptors also possess 7
			extracellular residues	
			Comment: Number of members:	preceding the cysteine box. 79
			ramber of members.	79
Acyl-ACP_TE	ŀ	Acyl-ACP thioesterase	Accession number:	PF01643
				cyl-ACP thioesterase
			Alignment method of	ashton M, Bateman A seed: Clustalw
			Source of seed memi	bers: Pfam-B_928 (release 4.1)
			Gathering cutoffs:	25 25
			Trusted cutoffs: Noise cutoffs:	91.70 91.70
				-192.80 -192.80 line: hmmbuild -F HMM SEED
			HMM build command	line: hmmcalibrateseed 0 HMM
			Reference Number:	[1]
			Reference Medline: Reference Title:	96068671
			an acyl-acyl	Modification of the substrate specificity of
			Reference Title:	carrier protein thioesterase by protein
			engineering.	
			Reference Author: Reference Location:	Yuan L, Voelker TA, Hawkins DJ; Proc Natl Acad Sci U S A 1995;92:10639-
			10643.	1700 Hall Adda 30, 0 3 A 1993,92.10039-
			Reference Number:	[2]
			Reference Medline: Reference Title:	92320297
			medium chains in	Fatty acid biosynthesis redirected to
			Reference Title:	transgenic oilseed plants.
			Reference Author:	Voelker TA, Worrell AC, Anderson L,
			Bleibaum J, Fan C, Reference Author:	Hawkins DJ, Radke SE, Davies HM;
	1		Reference Location:	Science 1992;257:72-74.
	ļ		Database Reference	INTERPRO; IPR002864;
1			Comment:	This family consists of various acyl-acyl
	İ		carrier protein (ACP) Comment:	thingsterages (TE) those terminate fathers
			group extension via	thioesterases (TE) these terminate fatty acyl
			Comment:	hydrolyzing an acyl group on a fatty acid [1].
			Number of members:	30
Acyltransferase		Acyltransferase	Accession number:	PF01553
				cyltransferase
			Author: Ba	teman A
			Alignment method of s	seed: Clustalw
			Gathering cutoffs:	ers: Pfam-B_128 (release 4.0) 8 8
			Trusted cutoffs:	14.40 14.40

Pfam	- In	I=	
	Prosite	Full Name	Description
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
		1	Reference Medline: 97411131 Reference Title: Barth syndrome may be due to an
			Reference Title: Barth syndrome may be due to an acyltransferase deficiency.
			Reference Author: Neuwald AF:
			Reference Location: Curr Biol 1997;7:465-466.
			Reference Number: [2]
			Reference Medline: 96224398
			Reference Title: A novel X-linked gene, G4.5. is responsible
			for Barth Reference Title: syndrome.
			Reference Title: syndrome. Reference Author: Bione S, D'Adamo P, Maestrini E,
			Gedeon AK, Bolhuis PA,
			Reference Author: Toniolo D;
			Reference Location: Nat Genet 1996;12:385-389.
			Database Reference INTERPRO; IPR002123;
			Database reference: PFAMB; PB009622;
			Database reference: PFAMB; PB009717;
			Database reference: PFAMB; PB033259; Database reference: PFAMB; PB041102:
			Database reference: PFAMB; PB041102; Database reference: PFAMB; PB041638;
			Comment: PPANIB; PB041638; Comment: This family contains acyltransferases
			involved in phospholipid
			Comment: biosynthesis and other proteins of unknown
	1 1		function [1]. This
			Comment: family also includes tafazzin Swiss:Q16635,
	. 1		the Barth syndrome Comment: gene [2]
			Comment: gene [2]. Number of members: 74
Adaptin_N		Adaptin N terminal region	Accession number: PF01602
			Definition: Adaptin N terminal region
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw Source of seed members: Pfam-B_491 (release 4.0)
			Gathering cutoffs: 12 12
			Trusted cutoffs: 15.50 15.50
			Noise cutoffs: 9.00 9.00
			HMM build command line: hmmbuild -f HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1] Reference Medline: 97409270

			Reference Title: Linking cargo to vesicle formation: receptor tail
			tail
			tail Reference Title: interactions with coat proteins. Reference Author: Kirchhausen T, Bonifacino JS, Riezman
			tail Reference Title: interactions with coat proteins. Reference Author: Kirchhausen T, Bonifacino JS, Riezman H;
			tail Reference Title: interactions with coat proteins. Reference Author: Kirchhausen T, Bonifacino JS, Riezman H; Reference Location: Curr Opin Cell Biol 1997;9:488-495.
			tail Reference Title: interactions with coat proteins. Reference Author: Kirchhausen T, Bonifacino JS, Riezman H; Reference Location: Curr Opin Cell Biol 1997;9:488-495. Reference Number: [2]
			tail Reference Title: interactions with coat proteins. Reference Author: Kirchhausen T, Bonifacino JS, Riezman H; Reference Location: Curr Opin Cell Biol 1997;9:488-495. Reference Medline: 89202379
			tail Reference Title: interactions with coat proteins. Reference Author: Kirchhausen T, Bonifacino JS, Riezman H; Reference Location: Curr Opin Cell Biol 1997;9:488-495. Reference Medline: 89202379
			tail Reference Title: interactions with coat proteins. Reference Author: H; Reference Location: Curr Opin Cell Biol 1997;9:488-495. Reference Medline: 89202379 Reference Title: Structural and functional division into two domains of the Reference Title: large (100- to 115-kDa)chains of the
			tail Reference Title: interactions with coat proteins. Reference Author: H; Reference Location: Curr Opin Cell Biol 1997;9:488-495. Reference Medline: 89202379 Reference Title: domains of the Reference Title: clathrin-associated interactions with coat proteins. Kirchhausen T, Bonifacino JS, Riezman Curr Opin Cell Biol 1997;9:488-495. [2] 89202379 Structural and functional division into two
			tail Reference Title: interactions with coat proteins. Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: domains of the Reference Title: clathrin-associated Reference Title: protein complex AP-2.
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: domains of the Reference Title: dotathrin-associated Reference Title: Reference Author: Reference Author: Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Title: Reference Author: Reference Author: Reference Title: Reference Author: Reference Title: Reference Author: Reference Title:
			tail Reference Title: interactions with coat proteins. Reference Author: H; Reference Location: Reference Number: [2] Reference Medline: Reference Title: domains of the Reference Title: clathrin-associated Reference Title: Reference Author: W, Vaisberg A, Chow interactions with coat proteins. Kirchhausen T, Bonifacino JS, Riezman Curr Opin Cell Biol 1997;9:488-495. [2] 89202379 Structural and functional division into two division into two divisions of the clathrin-associated Reference Title: Protein complex AP-2. RAKirchhausen T, Bonifacino JS, Riezman Curr Opin Cell Biol 1997;9:488-495. [2] 89202379 Structural and functional division into two divisions of the clathrin-associated Reference Title: Protein complex AP-2. RAKirchhausen T, Bonifacino JS, Riezman
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: domains of the Reference Title: clathrin-associated Reference Title: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Location: Reference Location: Reference Medline: Beg202379 Structural and functional division into two division into t
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: domains of the Reference Title: clathrin-associated Reference Title: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Location: Reference Location: Reference Author: W, Vaisberg A, Chow Reference Location: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Title: Referenc
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: Odomains of the Reference Title: Clathrin-associated Reference Title: Reference Author: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Author: Reference Location: Reference Author: Reference Author: Reference Author: Reference Author: Reference Location: Reference Author: Reference Location: Reference Author: Reference Location: Reference Author: Reference Author: Reference Location: Reference Author: Reference Author: Reference Location: Reference Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Reference Author: Reference Author: Reference Redline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Author: Reference Title: Reference
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: domains of the Reference Title: dotathrin-associated Reference Title: Reference Author: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Author: Reference Location: W, Vaisberg A, Chow Reference Author: Reference Location: 2616. Database Reference Database reference: Interactions with coat proteins. Kirchhausen T, Bonifacino JS, Riezman Curr Opin Cell Biol 1997;9:488-495. [2] 89202379 Structural and functional division into two large (100- to 115-kDa)chains of the complex AP-2. RAKirchhausen T, Nathanson KL, Matsui EP, Burne C, Keen JH, Davis AE; Proc Natl Acad Sci U S A 1989;86:2612- INTERPRO; IPR002553; PFAMB; PB040953;
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: domains of the Reference Title: dotathrin-associated Reference Author: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Author: Reference Location: W, Vaisberg A, Chow Reference Author: Reference Location: Reference Location: W, Vaisberg A, Chow Reference Author: Reference Location: Reference Location: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Title: Refere
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: domains of the Reference Title: dotathrin-associated Reference Title: Reference Author: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Author: Reference Location: W, Vaisberg A, Chow Reference Author: Reference Location: Reference Location: BREFEROE Author: Reference Title: Reference
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Medline: Reference Title: domains of the Reference Title: clathrin-associated Reference Title: Reference Author: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Location: Reference Author: Reference Title: Ref
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: Reference Title: Reference Title: Reference Title: Curr Opin Cell Biol 1997;9:488-495. [2] 89202379 Structural and functional division into two domains of the Reference Title: Clathrin-associated Reference Title: Reference Author: W, Vaisberg A, Chow Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Title: Intractions with coat proteins. Kirchhausen T, Bonifacino JS, Riezman 49202379 Structural and functional division into two domains of the clark of the complex AP-2. RAKirchhausen T, Bonifacino JS, Riezman 49202379 Structural and functional division into two domains of the clark of the complex AP-2. RAKirchhausen T, Nathanson KL, Matsui EP, Burne C, Keen JH, Davis AE; Proc Natl Acad Sci U S A 1989;86:2612- 2616. INTERPRO; IPR002553; PFAMB; PB040953; This family consists of the N terminal region of various alpha, Comment: beta and gamma subunits of the AP-1, AP-2 protein complexes. The adaptor proteins.
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Medline: Reference Title: domains of the Reference Title: clathrin-associated Reference Author: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Author: Reference Author: W, Vaisberg A, Chow Reference Location: 2616. Database Reference Database Reference Database reference: Comment: of various alpha, Comment: and AP-3 adaptor Comment: complexes are involved in
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Title: domains of the Reference Title: dotathrin-associated Reference Title: Reference Author: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Location: W, Vaisberg A, Chow Reference Location: Reference Location: W, Vaisberg A, Chow Reference Location: Conment: Database Reference Database reference: Comment: Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious alpha, Comment: Drawious with coat proteins. Kirchhausen T, Bonifacino JS, Riezman Curr Opin Cell Biol 1997;9:488-495. [2] 89202379 Structural and functional division into two darge (100- to 115-kDa)chains of the complex AP-2. RAKirchhausen T, Nathanson KL, Matsui EP, Burne C, Keen JH, Davis AE; Proc Natl Acad Sci U S A 1989;86:2612- Database reference Databas
			tail Reference Title: Reference Author: H; Reference Location: Reference Number: Reference Medline: Reference Medline: Reference Title: domains of the Reference Title: clathrin-associated Reference Author: W, Vaisberg A, Chow Reference Author: W, Vaisberg A, Chow Reference Author: Reference Author: W, Vaisberg A, Chow Reference Location: 2616. Database Reference Database Reference Database reference: Comment: of various alpha, Comment: and AP-3 adaptor Comment: complexes are involved in

		5	338
Pfam	Prosite	Full Name	Description
			Comment: by comparison to the C-terminal which is variable within members of the Comment: AP-2 family[2]; and it has been proposed that this constant region Comment: interacts with another uniform component of the coated vesicles [2]. Number of members: 66
ALAD	PDOC00153	Delta-aminolevulinic acid dehydratase active site	Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (ALAD) [1] catalyzes the second step in the biosynthesis of heme, the condensation of two molecules of 5-aminolevulinate to form porphobilinogen. The enzyme is an oligomer composed of eight identical subunits. Each of the subunits binds an atom of zinc or of magnesium (in plants). A lysine has been implicated in the catalytic mechanism [2]. The sequence of the region in the vicinity of the active site residue is conserved in ALAD from various prokaryotic and eukaryotic species.
			Description of pattern(s) and/or profile(s) Consensus pattern G-x-D-x-[LIVM](2)-[IV]-K-P-[GSA]-x(2)-Y [K is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Pattern and text revised. References [1] Li JM., Russell C.S., Cosloy S.D. Gene 75:177-184(1989). [2] Gibbs P.N.B., Jordan P.M. Biochem. J. 236:447-451(1986).
Aldolase	PDOC00144	KDPG and KHG aldolases active site signatures	4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) (KHG-aldolase) catalyzes the interconversion of 4-hydroxy-2-oxoglutarate into pyruvate and glyoxylate. Phospho-2-dehydro-3-deoxygluconate aldolase (EC 4.1.2.14) (KDPG-aldolase) catalyzes the interconversion of 6-phospho-2-dehydro-3-deoxy-D-gluconate into pyruvate and glyceraldehyde 3-phosphate. These two enzymes are structurally and functionally related [1]. They are both homotrimeric proteins of approximately 220 amino-acid residues. They are class I aldolases whose catalytic mechanism involves the formation of a Schiff-base intermediate between the substrate and the epsilon-amino group of a lysine residue. In both enzymes, an arginine is required for catalytic activity. We developed two signature patterns for these enzymes. The first one contains the active site arginine and the second, the lysine involved in the Schiff-base formation.
			Description of pattern(s) and/or profile(s)





			839
Pfam	Prosite	Full Name	Description Consensus pattern G-[LIVM]-x(3)-E-[LIV]-T-[LF]-R [R is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for Bacillus subtilis KDPG-aldolase which has Thr instead of Arg in the active site. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-x(3)-[LIVMF]-K-[LF]-F-P-[SA]-x(3)-G [K is involved in Schiff-base formation] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Vlahos C J., Dekker E.E. J. Biol. Chem. 263:11683-11691(1988).
Alpha_L_fucos	PDOC00324	Alpha-L-fucosidase	Alpha-L-fucosidase (EC 3.2.1.51) [1] is a lysosomal enzyme responsible for hydrolyzing the alpha-1,6-linked fucose joined to the reducing-end N-acetylglucosamine of the carbohydrate moieties of glycoproteins. Deficiency of alpha-L-fucosidase results in the lysosomal storage disease fucosidosis. A cysteine residue is important for the activity of the enzyme. There is only one cysteine conserved between the sequence of mammalian alpha-L-fucosidase and that of the slime mold Dictyostelium discoideum. We have derived a pattern from the region around that conserved cysteine. Description of pattern(s) and/or profile(s) Consensus pattern P-x(2)-L-x(3)-K-W-E-x-C [C is the putative active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note these proteins belong to family 29 in the classification of glycosyl hydrolases [2,E1]. Last update November 1997 / Pattern and text revised. References [1]
Amino_oxidase		Flavin containing amine oxidase	Fisher K.J., Aronson N.N. Jr. Biochem. J. 264:695-701(1989). [2] Henrissat B. Biochem. J. 280:309-316(1991). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt Accession number: PF01593 Definition: Flavin containing amine oxidase Author: Bashton M, Bateman A Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_606 (release 4.1) Gathering cutoffs: -110 -110 Trusted cutoffs: -110.00 -110.00 Noise cutoffs: -111.80 -111.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 98258926

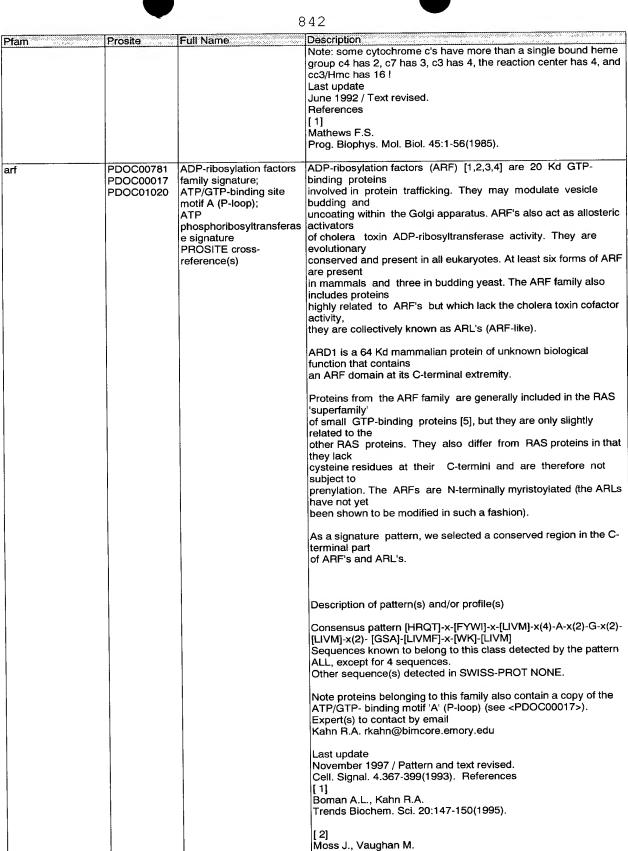




Pfam	Prosite	Full Name	Description	
			Reference Title:	Maize polyamine oxidase: primary structure
			from protein and	
			Reference Title:	cDNA sequencing.
			Reference Author:	Tavladoraki P, Schinina ME, Cecconi F,
			Agostino SD, Manera	
			Reference Author:	F, Rea G, Mariottini P, Federico R,
			Angelini R;	
			Reference Location:	FEBS Lett 1998;426:62-66.
1			Reference Number:	[2]
	1		Reference Medline:	97306298
	l		Reference Title:	A key amino acid responsible for substrate
			selectivity of	, , , , , , , , , , , , , , , , , , ,
			Reference Title:	monoamine oxidase A and B.
			Reference Author:	Tsugeno Y, Ito A;
			Reference Location:	J Biol Chem 1997;272:14033-14036.
			Reference Number:	[3]
			Reference Medline:	95287865
			Reference Title:	Cloning, sequencing and heterologous
			expression of the	Clothing, sequencing and heterologous
	-		Reference Title:	monoamine oxidase gene from Aspergillus
	-			monoamme oxidase gene nom Asperginus
	1		niger.	Schilling B. Lorch V.
			Reference Author:	Schilling B, Lerch K;
			Reference Location:	Mol Gen Genet 1995;247:430-438.
			Database Reference:	SCOP; 1b37; fa; [SCOP-USA][CATH-
			PDBSUM]	W.T.C.D.D.O. IDDA
			Database Reference	INTERPRO; IPR002937;
			Database Reference	PDB; 1b37 A; 14; 455;
			Database Reference	PDB; 1b5q A; 14; 455;
			Database Reference	PDB; 1b37 B; 14; 455;
	1		Database Reference	PDB; 1b37 C; 14; 455;
			Database Reference	PDB; 1b5q B; 14; 455;
			Database Reference	PDB; 1b5q C; 14; 455;
	:		Database reference:	PFAMB; PB017518;
			Database reference:	PFAMB; PB024839;
			Database reference:	PFAMB; PB040747;
			Comment:	This family consists of various amine
			oxidases, including m	•
	i			oxidase (PAO) [1] and various flavin
			containing monoamin	
		i		(MAO). The aligned region includes the
			flavin binding site of the	
			_	enzymes.
1				In vertebrates MAO plays an important role
			regulating the intracel	
				levels of amines via there oxidation; these
	1		include various	levels of armines via there extraction, these
			_	neurotransmitters incurotoxins and trace
			1	neurotransmitters, neurotoxins and trace
			amines [2]. In lower e Comment:	
				such as aspergillus and in bacteria the main
			role of amine oxidase	
				to provide a source of ammonium [3].
			-	PAOs in plants, bacteria and protozoa
			oxidase spermidine a	
				to an aminobutyral, diaminopropane and
1			hydrogen peroxide ar	
1				involved in the catabolism of polyamines [1].
				Other members of this family include
			tryptophan 2-monoox	
				putrescine oxidase, corticosteroid binding
			proteins and antibacte	
				glycoproteins.
			Number of members:	58
ANF_receptor	PDOC00430	Natriuretic peptides	Natriuretic peptides a	are hormones involved in the regulation of
		receptors signature	fluid and	
	-	_	electrolyte homeostas	sis. These hormones stimulate the
			intracellular productio	
			of cyclic GMP as a se	econd messenger.
			Currently, three types	of natriuretic peptide receptors are known
			[1,2]. Two	•
			1	clase activity: GC-A (or ANP-A) which
			seems specific to	• • • •
				tide (ANP), and GC-B (or ANP-B) which
·	.1	·		



		3	341
Pfam	Prosite	Full Name	Description
			seems to be stimulated more effectively by brain natriuretic peptide (BNP) than by ANP. The third receptor (ANP-C) is probably responsible for the clearance of ANP from the circulation and does not play a role in signal transduction.
			GC-A and GC-B are plasma membrane-bound proteins that share the following topology: an N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain followed by a large cytoplasmic C-terminal region that can be subdivided into two domains: a protein kinase-like domain (see <pdoc00100>) that appears important for proper signalling and a guanylate cyclase catalytic domain (see <pdoc00425>). The topology of ANP-C is different: like GC-A and -B it possesses an extracellular ligand-binding region and a transmembrane domain, but its cytoplasmic domain is very short. We developed a pattern from the ligand-binding region of natriuretic peptide receptors based on a highly conserved region located in the N-terminal part of the domain.</pdoc00425></pdoc00100>
			Description of pattern(s) and/or profile(s) Consensus pattern G-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update May 1991 / First entry. References [1] Garbers D.L. New Biol. 2:499-504(1990). [2] Schulz S., Chinkers M., Garbers D.L.
Apocytochrome_F	PDOC00169	Cytochrome c family heme-binding site signature	FASEB J. 2:2026-2035(1989). In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c.
			Description of pattern(s) and/or profile(s) Consensus pattern C-{CPWHF}-{CPWR}-C-H-{CFYW} Sequences known to belong to this class detected by the pattern ALL, except for four cytochrome c's which lack the first thioether bond. Other sequence(s) detected in SWISS-PROT 454.



Moss J., Vaughan M.





7-4-	T-2-10000		
Pfam	Prosite	Full Name	Description
	1		Prog. Nucleic Acid Res. Mol. Biol. 45:47-65(1993).
			[4]
			Amor J.C., Harrison D.H., Kahn R.A., Ringe D.
			Nature 372:704-708(1994).
İ			[5]
			Valencia A., Chardin P., Wittinghofer A., Sander C.
			Biochemistry 30:4637-4648(1991).
			From sequence comparisons and crystallographic data analysis it
			has been shown
			[1,2,3,4,5,6] that an appreciable proportion of proteins that bind
			ATP or GTP
			share a number of more or less conserved sequence motifs. The
			lest conserved of these motifs is a glycine-rich region, which typically forms a
			flexible
			loop between a beta-strand and an alpha-helix. This loop interacts
			with one of
			the phosphate groups of the nucleotide. This sequence motif
	, I		is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5].
	,		control to as the A consensus sequence [1] of the P-100p [5].
			There are numerous ATP- or GTP-binding proteins in which the
	J		P-loop is found.
,	,		We list below a number of protein families for which the
•	ļ		relevance of the presence of such motif has been noted:
			presence of such motif has been noted:
			- ATP synthase alpha and beta subunits (see <pdoc00137>).</pdoc00137>
			- Myosin heavy chains.
	İ		- Kinesin heavy chains and kinesin-like proteins (see
			<pdoc00343>). Dispersion and dispersion (like posterior (see PDOC00000)).</pdoc00343>
			- Dynamins and dynamin-like proteins (see <pdoc00362>) Guanylate kinase (see <pdoc00670>).</pdoc00670></pdoc00362>
			- Thymidine kinase (see <pdoc00524>).</pdoc00524>
			- Thymidylate kinase (see <pdoc01034>).</pdoc01034>
			- Shikimate kinase (see <pdoc00868>).</pdoc00868>
			- Nitrogenase iron protein family (nifH/frxC) (see <pdoc00580>).</pdoc00580>
			- ATP-binding proteins involved in 'active transport' (ABC transporters) [7]
			(see <pdoc00185>).</pdoc00185>
			- DNA and RNA helicases [8,9,10].
			- GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2,
			etc.).
			- Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.).
			- Nuclear protein ran (see <pdoc00859>).</pdoc00859>
			- ADP-ribosylation factors family (see <pdoc00781>).</pdoc00781>
			- Bacterial dnaA protein (see <pdoc00771>).</pdoc00771>
			- Bacterial recA protein (see <pdoc00131>).</pdoc00131>
			- Bacterial recF protein (see <pdoc00539>) Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt,</pdoc00539>
			G0, etc.).
			- DNA mismatch repair proteins mutS family (See
			<pdoc00388>).</pdoc00388>
	1		- Bacterial type II secretion system protein E (see
			<pdoc00567>).</pdoc00567>
			Not all ATP- or GTP-binding proteins are picked-up by this motif.
İ			A number of
			proteins escape detection because the structure of their ATP-
	1		binding site is
			completely different from that of the P-loop. Examples of such
			proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or
			GTP-binding
			proteins the flexible loop exists in a slightly different form; this is
			the
			case for tubulins or protein kinases. A special mention must be
	ĺ		reserved for
			adenylate kinase, in which there is a single deviation from the P-loop
			· ·



Q	1	1

pattern: in the last position Gly is found instead of Ser or Thr. Description of pattern(s) and/or profile(s) Consensus pattern (AG)-x(4)-G-K_(ST) Sequences known to belong to this class detected by the pattern a majority, Other sequence(s) detected in SWISS-PROT in addition to the proteins listed above, the "A motif is also for fund in a number of other proteins listed above, the "A motif is also for fund in a number of other proteins (active proteins probably bind a nucleotide, becample chymothypsin year) ATP or GTP-building (as for example the hymothypsin year). The proteins fund is a nucleotide, becample chymothypsin year and Pro GTP-building (as for example). Last update July 1999 / Text revised. In the control of the proteins of the proteins fund in the proteins fund				3 4 4
Description of pattern(s) and/or profile(s) Consensus pattern [AG]-x(s)-AK-IST] Sequences known to belong to this class detected by the pattern a majority. Other sequence(s) detected in SWISS-PROT in addition to the proteins listed above, the 'A' motif is also found in a number of other proteins. Most of these proteins probably bind a nucleotide, belonging the proteins and interesting the protein of the proteins in a nucleotide, belonging to contact by email and the protein of pattern of the protein of the protein of pattern of the protein of pattern of the protein of the protein of pattern of the protein of pattern of the protein of pattern of the protein of pattern of the pattern of the protein of pattern of the	Pfam	Prosite	Full Name	
Consensus pattern [AG]-x(4)-G-K-{ST] Sequences known to belong to this class detected by the pattern a majority. Other sequence(s) detected in SWISS-PROT in addition to the proteins proteins. Most of these proteins probably hid a nucleotide, but others are definitively not ATP- or GTP-binding (as for example chymotryps), or human ferritin light chain). Expert(s) to contact by email Konnin E.V. koonin@mcb.him.nih.gov Last update Jay 1999 / Text revised. Bull J. 1945-951(1982). [2] Moller W., Amons R. FEBS Latt. 1961-7(1985). [3] Fy D.C., Kuby S.A., Mildvan A.S. Froc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 83:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gili D.R., Gallagher M.P. J. Bioenerg, Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Niehi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nuciele Acids Ries. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step. in the biosynthesis of histidine in bacteria, fungi and plants: Is a protein of about 23 to 32 Kd. As a signature pattern we selected a region. Description of pattern(s) and/or profile(s) Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X				pattern: in the last position Gly is found instead of Ser or Thr.
Consensus pattern [AG]-x(4)-G-K-{ST] Sequences known to belong to this class detected by the pattern a majority. Other sequence(s) detected in SWISS-PROT in addition to the proteins proteins. Most of these proteins probably hid a nucleotide, but others are definitively not ATP- or GTP-binding (as for example chymotryps), or human ferritin light chain). Expert(s) to contact by email Konnin E.V. koonin@mcb.him.nih.gov Last update Jay 1999 / Text revised. Bull J. 1945-951(1982). [2] Moller W., Amons R. FEBS Latt. 1961-7(1985). [3] Fy D.C., Kuby S.A., Mildvan A.S. Froc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 83:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gili D.R., Gallagher M.P. J. Bioenerg, Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Niehi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nuciele Acids Ries. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step. in the biosynthesis of histidine in bacteria, fungi and plants: Is a protein of about 23 to 32 Kd. As a signature pattern we selected a region. Description of pattern(s) and/or profile(s) Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X				
Consensus pattern [AG]-x(4)-G-K-{ST] Sequences known to belong to this class detected by the pattern a majority. Other sequence(s) detected in SWISS-PROT in addition to the proteins proteins. Most of these proteins probably hid a nucleotide, but others are definitively not ATP- or GTP-binding (as for example chymotryps), or human ferritin light chain). Expert(s) to contact by email Konnin E.V. koonin@mcb.him.nih.gov Last update Jay 1999 / Text revised. Bull J. 1945-951(1982). [2] Moller W., Amons R. FEBS Latt. 1961-7(1985). [3] Fy D.C., Kuby S.A., Mildvan A.S. Froc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 83:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gili D.R., Gallagher M.P. J. Bioenerg, Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Niehi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nuciele Acids Ries. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step. in the biosynthesis of histidine in bacteria, fungi and plants: Is a protein of about 23 to 32 Kd. As a signature pattern we selected a region. Description of pattern(s) and/or profile(s) Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X				
Consensus pattern [AG]-x(4)-G-K-{ST] Sequences known to belong to this class detected by the pattern a majority. Other sequence(s) detected in SWISS-PROT in addition to the proteins proteins. Most of these proteins probably hid a nucleotide, but others are definitively not ATP- or GTP-binding (as for example chymotryps), or human ferritin light chain). Expert(s) to contact by email Konnin E.V. koonin@mcb.him.nih.gov Last update Jay 1999 / Text revised. Bull J. 1945-951(1982). [2] Moller W., Amons R. FEBS Latt. 1961-7(1985). [3] Fy D.C., Kuby S.A., Mildvan A.S. Froc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 83:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gili D.R., Gallagher M.P. J. Bioenerg, Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Niehi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nuciele Acids Ries. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step. in the biosynthesis of histidine in bacteria, fungi and plants: Is a protein of about 23 to 32 Kd. As a signature pattern we selected a region. Description of pattern(s) and/or profile(s) Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-Consensus pattern E.X(5)-G-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X-S(AG) (X2) [W]-X-D-[LV]-X(2)-(ST)-X				Description of pattern(s) and/or profile(s)
Sequences known to belong to this class detected by the pattern a majority. Other sequence(s) detected in SWISS-PROT in addition to the proteins listed above, the 'A' motif is also found in a number of other proteins. Most of these proteins probably bind a nucleotide, before the proteins and the proteins of the prot				, , , , , , , , , , , , , , , , , , , ,
majority. Other sequence(s) detected in SWISS-PROT in addition to the proteins listed above, the "A" motif is also found in a number of other proteins in the proteins probably bind a nucleotide, but of the proteins probably bind a nucleotide, but of the proteins probably bind a nucleotide, but of the proteins probably bind a nucleotide, but of the protein proteins protein proteins protein proteins protein proteins protein proteins protein protein proteins protein pr				Consensus pattern [AG]-x(4)-G-K-[ST]
Other sequence(s) detected in SWISS-PROT in addition to the proteins listed above, the 'A' modif is also found in a number of other proteins. Most of these proteins probably bind a nucleotide, but others are definitively not ATP or GTP-binding (ast for example chymotrypsin, or human ferritin light chain). Expert(s) to contact by email Koonin E.V. Koonin@ncbl.nlm.nlh.gov Last update July 1999 / Text revised. References [1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982). [2] Moller W., Amons R., FESS Last. 1861-7(1985). [3] Fy D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83-907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 83-907-911(1986). [5] Saraste M., Sibbald P.R., Wiltinghoter A. Trends Blochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bloenerg, Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Lercy P., Nielsen P.J., Nishi K., Schnier J., Slommaki P.P. Nature 337:121-12(1989). 10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E.×(C)-G.×(EAG.) ×(C)-[V]×-D-(LUV)-X(2)-(ST)-VIII-VIII-VIII-VIII-VIII-VIII-VIII-VI				
proteins listed above, the 'A' motif is also found in a number of other proteins probably bird a nucleotide, but others are definitively not ATP- or GTP-binding (as for example chymotrypsin, or human ferrifful light chain). Expert(s) to contact by email (Aconin E.V. koonin@ncbi.nlm.nlh.gov) Last update July 1999 / Text revised. References 11 Walker J.E., Saraste M., Runswick M.J., Gay N.J. [EMBO J. 1:945-951(1982). [2] Moller W., Amons R. [EBS Lett. 1861-7(1985). [3] Fy D.C., Kuby S.A., Mildvan A.S. [2] Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. [2] Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Blochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). [41] ATP phosphoribosyltransferase (EC 2-4.2-17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants: 1. Is a protein of about 23 to 32 Kd. As a signature pattern we seeded an engline me. Excipilization Excipiling and plants: 1. Is a protein of about 23 to 32 Kd. As a signature pattern we seeded an engline me. Excipiling and plants: 1. Is a protein of about 23 to 32 Kd. As a signature pattern we seeded an engline me. Excipiling and plants: 1. Is a protein of about 23 to 32 Kd. As a signature pattern we seeded an engline me. Excipiling and plants: 1. Is a protein of pattern(s) and/or profile(s) Consensus pattern Excipiling and plants: 1. Is a protein of pattern(s) and/or profile(s) Consensus pattern Excipiling and plants: 1. Is a protein of pattern(s) and/or profile(s) Consensus pattern Excipiling and plants: 1. Is a protein of a signature pattern we seede				
but others are definitively not ATP- or GTP-binding (as for example chymotrypsin, or human ferrifn light chain). Expert(s) to contact by email Noonin E.V. koonin E.V. koonin E.V. koonin@ncbi.nlm.nih.gov Last update July 1999/Text revised. References 11 11 Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982). [2] Moller W., Amons R. FEBS Lett. 186:1-7(1985). [3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plaints. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profilie(s).				proteins listed above, the 'A' motif is also found in a number of
example chymotrypsin, or human ferritin light chain). Expert(s) to contact by email Koonin E.V. koonin@ncbl.nlm.nih.gov Last update July 1999 / Text revised. References [Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982). [2] Moller W., Amons R., FEBS Lett. 186:1-7(1985). [3] Fry D.C., Kuby S.A., Mildvan A.S., Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C., Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Blochem. Sci. 15:430-434(1990). [6] Koonin E.V., J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Minmack M.M., Gileadi U., Gill D.R., Gailagher M.P., J. Bloenerg, Biomembr. 22:571-592(1990). [8] Hodgman T.C., Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P., Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It also also spattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Description of pattern(s) and/or profilie(s) Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensus pattern Exp(5):G.x.(SAG)-x(2)-[IV]-x-D-(LIV]-x(2)-(ST)-Consensu	i			other proteins. Most of these proteins probably bind a nucleotide,
Expert(s) to contact by email Koonin E.V. koonin E.V. koonin(A) Koonin(A				
Last update July 1999 / Text revised. References [1] Walker J.E., Saraste M., Runawick M.J., Gay N.J. EMBO. J.1945-951(1982). [2] Moller W., Amons R. FEBS Lett. 186:1-7(1985). [3] Fy D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghoffer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Höggins C.F., Hyde S.C., Mirmmack M.M., Gileadi U., Gili D.R., Gallagher M.P. J. Bioenerg, Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Sionimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res., 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern Ex.(5):G.x.(SAG)+(2)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)+(2)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)+(2)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)+(2)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)+(2)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)+(2)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)+(2)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)-(x)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)-(x)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)-(x)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)-(x)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)-(x)-[U/]-x-D-(LIV]-x(2)-[ST]-Consensus pattern Ex.(5):G.x.(SAG)-(x)-[U/				Expert(s) to contact by email
July 1999 / Text revised. References [1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982). [2] Moller W., Amons R. FEBS Lett. 186:1-7(1985). [3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Blochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Blol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bloenerg. Blomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi X., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungl and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(e) and/or profile(s)				Koonin E.V. koonin@ncbi.nlm.nih.gov
July 1999 / Text revised. References [1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982). [2] Moller W., Amons R. FEBS Lett. 186:1-7(1985). [3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Blochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Blol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bloenerg. Blomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi X., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungl and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(e) and/or profile(s)				Last undate
References				
Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982). [2] Moller W., Amons R. FEBS Lett. 186:1-7(1985). [3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mirmmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Sjonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E.x(5)-G-x-(SAG)-x(2)-[IV]-x-D-[LIV]-x(2)-(ST)-		İ		
EMBO J. 1:945-951(1982). [2] [2] [3] Fy D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Sionimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-(SAG)-x(2)-[IV]-x-D-[LIV]-x(2)-(ST]-				
[2] Moller W., Amons R. FEBS Lett. 186:1-7(1985). [3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s)				
Möller W., Amons R. FEBS Lett. 186:1-7(1985). [3] Fy D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Blochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mirmmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schmier J., Slominski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or protile(s)				Z. 1.575-55 (1562).
FEBS Lett. 186:1-7(1985). [3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or protile(s)				
[3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Blochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Sionimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s)				
Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s)				FEDS Lett. 1880;1-7(1980).
Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s)				[3]
[4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s)				Fry D.C., Kuby S.A., Mildvan A.S.
Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mirmmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E.x(5)-G.x-[SAG]-x(2)-[V]-x-D-[LIV]-x(2)-[ST]-				Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986).
Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mirmmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E.x(5)-G.x-[SAG]-x(2)-[V]-x-D-[LIV]-x(2)-[ST]-			•	[4]
Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-{SAG}-x(2)-[IV]-x-D-[LIV]-x(2)-{ST}-				Dever T.E., Glynias M.J., Merrick W.C.
Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).
Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				f 51
Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mirnmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-		İ	ľ	Trends Biochem. Sci. 15:430-434(1990).
Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				(e)
[7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E.x(5)-Gx-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-	i		į.	J. Mol. Biol. 229:1165-1174(1993).
Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E.x(5)-Gx-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				[7]
Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
[8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-			į.	Gallagher M.P.
Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-		-	•	J. Bioenerg. Biomembr. 22:571-592(1990).
Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				[8]
Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-			Į:	Hodgman T.C.
Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				[9]
K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
[10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-]1	K., Schnier J., Slonimski P.P.
Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-			ļ!	Nature 337:121-122(1989).
Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-			ا	1101
Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-		j		ATP phosphoribosyltransforase (EC 2.4.2.17) in the one was that
first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-		-	1	first step in the biosynthesis of histidine in bacteria, fungi and
selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-				
Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]- G-x-T-[LM]			Į	Description of pattern(s) and/or profile(s)
G-x-T-ſLM1			lo	Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[STI-
A CONTRACT OF THE PROPERTY OF		<u>l</u> ,		G-x-T-[LM]



	- 4
	- 1

Pfam	Prosite Full Name	
		Sequences known to belong to this class detected by the pattern ALL.
		Other sequence(s) detected in SWISS-PROT NONE.
		Last update
		July 1998 / First entry.
		, , , , , , , , , , , , , , , , , , , ,
ArgJ	ArgJ family	Accession number: PF01960
		Definition: ArgJ family
		Author: Enright A, Ouzounis C, Bateman A
		Alignment method of seed: Clustalw
		Source of seed members: Enright A
		Gathering cutoffs: 25 25 Trusted cutoffs: 258.70 99.60
		Noise cutoffs: 7.10 7.10
		HMM build command line: hmmbuild -f HMM SEED
		HMM build command line: hmmcalibrateseed 0 HMM
		Reference Number: [1]
		Reference Medline: 93232760
		Reference Title: Primary structure, partial purification and
		regulation of
		Reference Title: key enzymes of the acetyl cycle of arginine
		biosynthesis in
		Reference Title: Bacillus stearothermophilus: dual function of ornithine
		Reference Title: acetyltransferase.
		Reference Author: Sakanyan V, Charlier D, Legrain C,
		Kochikyan A, Mett I,
		Reference Author: Pierard A, Glansdorff N;
		Reference Location: J Gen Microbiol 1993;139:393-402.
		Database Reference INTERPRO; IPR002813;
		Comment: Members of the ArgJ family catalyse the first
	•	EC:2.3.1.35 and Comment: fifth steps EC:2.3.1.1 in arginine
		biosynthesis.
		Number of members: 22
Armadillo_seg	Armadillo/beta-catenin-	Accession number: PF00514
	like repeats	Definition: Armadillo/beta-catenin-like repeats
		Author: Bateman A, Chris Ponting, Joerg Schultz, Peer
		Bork Alignment method of seed: Manual
		Source of seed members: SMART
		Source of seed members: SMART Gathering cutoffs: 24 0
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20
		Gathering cutoffs: 24.0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED
		Gathering cutoffs: 24.0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350
		Gathering cutoffs: 24.0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region
		Gathering cutoffs: 24.0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Location: Cell 1997;90:871-882.
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2]
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Medline: 96107551
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Medline: 96107551 Reference Title: Signal transduction of beta-catenin.
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Number: [2] Reference Medline: 96107551 Reference Author: Gumbiner BM;
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillor repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Number: [2] Reference Medline: 96107551 Reference Author: Gumbiner BM; Reference Location: Curr Opin Cell Biol 1995;7:634-640.
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Number: [2] Reference Medline: 96107551 Reference Author: Gumbiner BM;
		Gathering cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Medline: 96107551 Reference Author: Gumbiner BM; Reference Location: Gumbiner BM; Curr Opin Cell Biol 1995;7:634-640.
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillor repeat region Reference Title: Of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Medline: 96107551 Reference Author: Gumbiner BM; Reference Location: Gumbiner BM; Reference Number: [3] Reference Medline: 97454713 Armadillo and dTCF: a marriage made in the nucleus.
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Author: Reference Author: Reference Number: [2] Reference Medline: 96107551 Reference Author: Reference Location: Gumbiner BM; Reference Number: Reference Number: Reference Number: Reference Number: Reference Medline: Reference Title: Armadillo and dTCF: a marriage made in the nucleus. Reference Author: Cavallo R, Rubenstein D, Peifer M;
		Gathering cutoffs: 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Medline: 96107551 Reference Author: Gumbiner BM; Reference Location: Gumbiner BM; Reference Number: [3] Reference Medline: 97454713 Reference Medline: 97454713 Reference Author: Reference Author: Cavallo R, Rubenstein D, Peifer M; Reference Location: Curr Opin Genet Dev 1997;7:459-466.
		Gathering cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillor repeat region Reference Title: of beta-catenin. Reference Author: Reference Location: Cell 1997;90:871-882. [2] Reference Medline: 96107551 Reference Author: Gumbiner BM; Reference Location: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Cavallo R, Rubenstein D, Peifer M; Curr Opin Genet Dev 1997;7:459-466.
		Gathering cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Number: Reference Medline: Signal transduction of beta-catenin. Reference Author: Gumbiner BM; Reference Number: Reference Medline: 97454713 Reference Title: Armadillo and dTCF: a marriage made in the nucleus. Reference Author: Cavallo R, Rubenstein D, Peifer M; Reference Number: Reference Number: [4] Reference Medline: 94082295
		Gathering cutoffs: 24.0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillor repeat region Reference Title: of beta-catenin. Reference Author: Reference Number: [2] Reference Number: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Number: Reference Medline: Reference Medline: Reference Medline: Reference Author: Reference Author: Reference Medline: Reference Author: Reference Author: Reference Medline: Reference Author: Reference Aut
		Gathering cutoffs: 24.0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillor repeat region Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Medline: 96107551 Reference Author: Gumbiner BM; Reference Author: Reference Number: [3] Reference Number: Reference Number: Reference Author: Curr Opin Cell Biol 1995;7:634-640. Reference Title: Armadillo and dTCF: a marriage made in the nucleus. Reference Author: Cavallo R, Rubenstein D, Peifer M; Reference Number: Reference Number: Reference Medline: P4082295 Reference Title: Association of the APC tumor suppressor
		Gathering cutoffs: 24.0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: [1] Reference Number: [1] Reference Title: Three-dimensional structure of the armadillor repeat region Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Medline: 96107551 Reference Author: Gumbiner BM; Reference Location: Curr Opin Cell Biol 1995;7:634-640. Reference Number: [3] Reference Number: Reference Author: Cavallo R, Rubenstein D, Peifer M; Reference Number: Reference Numb
		Gathering cutoffs: 24.0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Author: Reference Author: Reference Number: [2] Reference Medline: 96107551 Reference Medline: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Medline: Reference Medline: Reference Medline: Reference Author: Reference Medline: Reference Number: Reference Number: Reference Number: Reference Number: Reference Medline: Reference Number: R
		Gathering cutoffs: 24.0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: [1] Reference Title: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Number: [2] Reference Medline: 96107551 Reference Author: Gumbiner BM; Reference Location: Curr Opin Cell Biol 1995;7:634-640. Reference Number: [3] 97454713 Armadillo and dTCF: a marriage made in the nucleus. Reference Author: Cavallo R, Rubenstein D, Peifer M; Reference Number: Reference Number: Reference Number: Reference Medline: Reference Medline: Reference Title: Association of the APC tumor suppressor protein with Reference Title: catenins.

		C	346	
Pfam	Prosite	Full Name	Description	ni necessaria (necessaria de la constanta de la constanta de la constanta de la constanta de la constanta de l
			Reference Title:	Association of the APC gene product with
			beta-catenin.	
			Reference Author:	Rubinfeld B, Souza B, Albert I, Muller O,
			Chamberlain SH,	Maria ED Maria C S 111 5
			Reference Author:	Masiarz FR, Munemitsu S, Polakis P;
			Reference Location:	Science 1993;262:1731-1734.
			Reference Number:	[6]
			Reference Medline:	91084846
			Reference Title: encodes a functionall	The segment polarity gene armadillo
			Reference Title:	modular protein that is the Drosophila
			homolog of human	modular protein that is the Drosophila
			Reference Title:	plakoglobin.
			Reference Author:	Peifer M, Wieschaus E;
		i	Reference Location:	Cell 1990;63:1167-1176.
			Database Reference:	SCOP; 3bct; fa; [SCOP-USA][CATH-
			PDBSUM]	
			Database Reference:	EXPERT; Chris.Ponting@human-
			anatomy.oxford.ac.uk	
			Database reference:	SMART; ARM;
			Database Reference Database Reference	INTERPRO; IPR000225;
			Database Reference	PDB; 1ee5 A; 417; 457;
			Database Reference	PDB; 1bk5 A; 417; 457; PDB; 1bk5 B; 417; 457;
			Database Reference	PDB; 1bk6 A; 417; 457;
			Database Reference	PDB; 1bk6 B; 417; 457;
			Database Reference	PDB; 1ee4 A; 417; 457;
			Database Reference	PDB; 1ee4 B; 417; 457;
			Database Reference	PDB; 1ejl l; 409; 449;
			Database Reference	PDB; 1ejy I; 409; 449;
			Database Reference	PDB; 1ial A; 409; 449;
			Database Reference	PDB; 1ee5 A; 246; 286;
			Database Reference	PDB; 1bk5 A; 246; 286;
			Database Reference	PDB; 1bk5 B; 246; 286;
			Database Reference Database Reference	PDB; 1bk6 A; 246; 286; PDB; 1bk6 B; 246; 286;
			Database Reference	PDB; 16k6 B; 246; 286; PDB; 1ee4 A; 246; 286;
			Database Reference	PDB; 1ee4 B; 246; 286;
			Database Reference	PDB; 1ejl l; 241; 280;
			Database Reference	PDB; 1ejy l; 241; 280;
			Database Reference	PDB; 1ial A; 241; 280;
			Database Reference	PDB; 1ee5 A; 288; 328;
			Database Reference	PDB; 1bk5 A; 288; 328;
			Database Reference	PDB; 1bk5 B; 288; 328;
			Database Reference	PDB; 1bk6 A; 288; 328;
			Database Reference Database Reference	PDB; 1064 A; 288; 328;
			Database Reference	PDB; 1ee4 A; 288; 328; PDB; 1ee4 B; 288; 328;
			Database Reference	PDB; 1ejl I; 282; 322;
			Database Reference	PDB; 1ejy I; 282; 322;
			Database Reference	PDB; 1ial A; 282; 322;
			Database Reference	PDB; 1ejl l; 151; 191;
			Database Reference	PDB; 1ejy l; 151; 191;
			Database Reference	PDB; 1ial A; 151; 191;
			Database Reference	PDB; 1ee5 A; 162; 202;
			Database Reference	PDB; 1bk5 A; 162; 202;
			Database Reference	PDB; 1bk5 B; 162; 202;
			Database Reference	PDB; 1bk6 A; 162; 202;
			Database Reference	PDB; 1bk6 B; 162; 202;
			Database Reference Database Reference	PDB; 1ee4 A; 162; 202;
			Database Reference	PDB; 1ee4 B; 162; 202; PDB; 1ee5 A; 330; 370;
			Database Reference	PDB; 1663 A; 330; 370; PDB; 16k5 A; 330; 370;
			Database Reference	PDB; 1bk5 B; 330; 370;
			Database Reference	PDB; 1bk6 A; 330; 370;
			Database Reference	PDB; 1bk6 B; 330; 370;
			Database Reference	PDB; 1ee4 A; 330; 370;
			Database Reference	PDB; 1ee4 B; 330; 370;
			Database Reference	PDB; 1ejl l; 324; 364;
			Database Reference	PDB; 1ejy I; 324; 364;
			Database Reference	PDB; 1 ial A; 324; 364;
			Database Reference	PDB; 1ee5 A; 372; 412;
		1	Database Reference	PDB; 1bk5 A; 372; 412;
			Database Reference	PDB; 1bk5 B; 372; 412;



			47
Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1bk6 A; 372; 412;
			Database Reference PDB; 1bk6 B; 372; 412;
			Database Reference PDB; 1ee4 A; 372; 412;
			Database Reference PDB; 1ee4 B; 372; 412;
			Database Reference PDB; 1ejl !; 366; 406;
			Database Reference PDB; 1ejy I; 366; 406;
			Database Reference PDB; 1al A; 366; 406;
			Database Reference PDB; 1ejl I; 108; 149;
			Database Reference PDB; 1ejy I; 108; 149; PDB; 1ial A; 108; 149;
			Database Reference PDB; 1ee5 A; 119; 160;
			Database Reference PDB; 1bk5 A; 119; 160;
			Database Reference PDB; 1bk5 B; 119; 160;
			Database Reference PDB; 1bk6 A; 119; 160;
			Database Reference PDB; 1bk6 B; 119; 160;
			Database Reference PDB; 1ee4 A; 119; 160;
			Database Reference PDB; 1ee4 B; 119; 160;
			Database Reference PDB; 3bct; 583; 623;
			Database Reference PDB; 2bct; 583; 623;
			Database Reference PDB; 3bct; 391; 429;
			Database Reference PDB; 2bct; 391; 429;
			Database Reference PDB; 3bct; 224; 264;
			Database Reference PDB; 2bct; 224; 264;
			Database Reference PDB; 3bct; 431; 473;
			Database Reference PDB; 2bct; 431; 473;
			Database Reference PDB; 3bct; 350; 390;
			Database Reference PDB; 2bct; 350; 390;
			Database Reference PDB; 1ejl l; 193; 238;
			Database Reference PDB; 1ejy l; 193; 238; PDB; 1ial A; 193; 238; PDB; 1ial A; 193; 238;
			Database Reference PDB; 1ee5 A; 204; 244;
			Database Reference PDB; 1bk5 A; 204; 244;
		1	Database Reference PDB; 1bk5 B; 204; 244;
		•	Database Reference PDB; 1bk6 A; 204; 244;
			Database Reference PDB; 1bk6 B; 204; 244;
			Database Reference PDB; 1ee4 A; 204; 244;
		l	Database Reference PDB; 1ee4 B; 204; 244;
			Database Reference PDB; 1ibr D; 399; 437;
	i		Database Reference PDB; 1ibr B; 399; 437;
			Database Reference PDB; 1qgk A; 399; 437;
			Database Reference PDB; 1qgr A; 399; 437;
			Database reference: PFAMB; PB002221;
			Database reference: PFAMB; PB002617;
			Database reference: PFAMB; PB004638; Database reference: PFAMB; PB012310;
			Database reference: PFAMB; PB012310; Database reference: PFAMB; PB040528;
		1	Database reference: PFAMB; PB041028;
			Comment: Approx. 40 amino acid repeat. Tandem
			repeats form super-helix of helices
			Comment: that is proposed to mediate interaction of
			beta-catenin with its ligands.
			Comment: CAUTION: This family does not contain all
	}		known armadillo repeats.
			Number of members: 597
ATP_synt_B_c	PDOC00137	ATP synthase alpha and	ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2]
	1	beta subunits signature	is a component of the cytoplasmic membrane of eubacteria, the inner membrane
			of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase
			complex is composed of
			an oligomeric transmembrane sector, called CF(0), and a catalytic
	1		core, called
			coupling factor CF(1). The former acts as a proton channel; the
			latter is
			composed of five subunits, alpha, beta, gamma, delta and
			epsilon. The
		1	sequences of subunits alpha and beta are related and both
			contain a
			nucleotide-binding site for ATP and ADP. The beta chain has
			catalytic
			activity, while the alpha chain is a regulatory subunit.
			Vacuolar ATPases [3] (V-ATPases) are responsible for acidifying
		<u> </u>	



	8	48
Pfam Prosite	Full Name	Description
ridiii i i joons		a variety of intracellular compartments in eukaryotic cells. Like F-ATPases,
		they are oligomeric complexes of a transmembrane and a catalytic
		sector. The sequence of the largest subunit of the catalytic sector (70 Kd) is related to
		that of F-ATPase beta subunit, while a 60 Kd subunit, from the same
		sector, is related to the F-ATPases alpha subunit [4].
		Archaebacterial membrane-associated ATPases are composed of three subunits.
		The alpha chain is related to F-ATPases beta chain and the beta chain is related to F-ATPases alpha chain [4].
		A protein highly similar to F-ATPase beta subunits is found [5]
		in some bacterial apparatus involved in a specialized protein export
		pathway that proceeds without signal peptide cleavage. This protein is
		known as fill in Bacillus and Salmonella, Spa47 (mxiB) in Shigella flexneri,
		HrpB6 in Xanthomonas campestris and yscN in Yersinia virulence
		plasmids.
		In order to detect these ATPase subunits, we took a segment of ten amino-acid
		residues, containing two conserved serines, as a signature pattern. The first
		serine seems to be important for catalysis - in the ATPase alpha chain at
		least - as its mutagenesis causes catalytic impairment.
		Description of pattern(s) and/or profile(s)
		Consensus pattern P-[SAP]-[LIV]-[DNH]-x(3)-S-x-S [The first S is
		a putative active site residue] Sequences known to belong to this class detected by the pattern ALL, except for the archaebacterium Sulfolobus acidocaldarius ATPase alpha chain where the first Ser is replaced by Gly. Other sequence(s) detected in SWISS-PROT 37.
		Note F-ATPase alpha and beta subunits, V-ATPase 70 Kd subuni and the archaebacterial ATPase alpha subunit also contain a copy of the ATP-binding motifs A and B (see <pdoc00017>).</pdoc00017>
		Last update November 1997 / Pattern and text revised.
		References
		[1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989).
		[2] Senior A.E. Physiol. Rev. 68:177-231(1988).
		[3] Nelson N. J. Bioenerg. Biomembr. 21:553-571(1989).
		[4] Gogarten J.P., Kibak H., Dittrich P., Taiz L., Bowman E.J., Bowman B.J., Manolson M.F., Poole R.J., Date T., Oshima T., Konishi J., Denda K., Yoshida M. Proc. Natl. Acad. Sci. U.S.A. 86:6661-665(1989).
		[5] Dreyfus G., Williams A.W., Kawagishi I., MacNab R.M. J. Bacteriol. 175:3131-3138(1993).



		8	349
Pfam	Prosite	Full Name	Description
ATP-gua_Ptrans	PDOC00103	ATP:guanido	ATP:guanido phosphotransferases are a family of structurally and
		phosphotransferases active site	functionally related enzymes [1,2] that reversibly catalyze the transfer of phosphate
			between ATP and various phosphogens. The enzymes that belongs to this family
			are:
			- Creatine kinase (EC 2.7.3.2) (CK) [3,4], which plays an important role in
			energy metabolism of vertebrates. It catalyzes the reversible transfer of
			high energy phosphate from ATP to creatine, generating phosphocreatine and
			ADP. There are at least four different, but very closely related, forms of
		į	CK. Two of the CK isozymes are cytosolic: the M (muscle) and B (brain)
			forms while the two others are mitochondrial. In sea urchin there is a flagellar isozyme, which consists of the triplication of a CK-
			domain Glycocyamine kinase (EC 2.7.3.1) (guanidoacetate kinase), an
			enzyme that catalyzes the transfer of phosphate from ATP to guanidoacetate.
			- Arginine kinase (EC 2.7.3.3), an enzyme that catalyzes the transfer of
			phosphate from ATP to arginine Taurocyamine kinase (EC 2.7.3.4), an annelid-specific enzyme
			that catalyzes the transfer of phosphate from ATP to taurocyamine.
			- Lombricine kinase (EC 2.7.3.5), an annelid-specific enzyme that catalyzes
			the transfer of phosphate from ATP to lombricine Smc74 [1], a cercaria-specific enzyme from Schistosoma
			mansoni. This enzyme consists of two CK-related duplicated domains. The substrate(s) specificity
			of Smc74 is not yet known.
			A cysteine residue is implicated in the catalytic activity of these enzymes.
			The region around this active site residue is highly conserved and can be used
			as a signature pattern.
			Description of pattern(s) and/or profile(s)
			Consensus pattern C-P-x(0,1)-[ST]-N-[IL]-G-T [C is the active site residue]
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update November 1997 / Pattern and text revised.
			References [1]
			Stein L.D., Harn D.A., David J.R. J. Biol. Chem. 265:6582-6588(1990).
			[2]
			Strong S.J., Ellington W.R. Biochim. Biophys. Acta 1246:197-200(1995).
			[3] Bessman SP., Carpenter C.L.
			Annu. Rev. Biochem. 54:831-862(1985).
			[4] Haas R.C., Strauss A.W. J. Biol. Chem. 265:6921-6927(1990).
ATP-synt D		ATP synthase subunit [Accession number: PF01813





Pfam Prosite Full Name Description Definition: ATP synthase suburant Author: Bashton M, Batema Alignment method of seed: Clustalw	
Author: Bashton M, Batema	
Alignment method of seed: Cilistaly	n A
	04 (rologgo 4.2)
Source of seed members: Pfam-B_13 Gathering cutoffs: 25 25	04 (release 4.2)
Trusted cutoffs: 157.80 157.80	
Noise cutoffs: -79.90 -79.90	
HMM build command line: hmmbuild -	F HMM SEED
HMM build command line: hmmcalibra	ateseed 0 HMM
Reference Number: [1]	
Reference Medline: 96324968	
Reference Title: Subunit structure genes of the A1A0	and organization of the
	Archaeon Methanosarcina
mazei Go1.	, , , , , , , , , , , , , , , , , , ,
	erg C, Wegerle E, Meier I,
Mayer F, Muller V;	
Reference Location: J Biol Chem 19	996;271:18843-18852.
Reference Number: [2]	
Reference Medline: 95132627	
	and a yeast gene (VMA8)
encoding the subunit Reference Title: D of the vacuola	r H(+)-ATPase
	diyan S, Nelson N;
	d Sci U S A 1995;92:497-
501.	
Database Reference INTERPRO;	IPR002699;
	f subunit D form various
ATP synthases	L. transporting and No.
Comment: including V-type I dependent.	H+ transporting and Na+
	gested to be an integral
part of the	,
	f the V-ATPase [2].
Number of members: 21	
ATZ TRZ Chlorohydrolase Accession number: PF01685	
ATZ_TRZ Chlorohydrolase Accession number: PF01685 Definition: Chlorohydrolase	
Author: Bashton M, Batemai	n A
Alignment method of seed: Clustalw	
Source of seed members: Pfam-B 119	92 (release 4.1)
Gathering cutoffs: -84 -84 Trusted cutoffs: -74.80 -74.80	
Noise cutoffs: -94.30 -94.30	
HMM build command line: hmmbuild -	F HMM SEED
HMM build command line: hmmcalibra	iteseed 0 HMM
Reference Number: [1]	
Reference Medline: 96326334	
Reference Title: Atrazine chlorohy Pseudomonas sp. strain ADP:	GIOIASE IIOIII
	enzyme purification, and
protein	• • • • • • • • • • • • • • • • • • • •
	published erratum appears
in J Bacteriol	0051
Reference Title: 1999 Jan;181(2):	
	Sadowsky MJ, Wackett LP; 96;178:4894-4900.
Reference Location. 3 Bacterior 198	70, 17 0.4034-4300.
Reference Medline: 96011356	
	ession of the s-triazine
hydrolase gene	
Deference Title: Arm A\ from Dhad	ococcus corallinus and
	combinant etraine canable
development of	Johnsmant Strains Capable
development of Reference Title: Rhodococcus rec	
development of Reference Title: Rhodococcus rec of dealkylating and	e herbicide atrazine.
development of Reference Title: Rhodococcus rec of dealkylating and Reference Title: dechlorinating the	e herbicide atrazine. ens W, Mulbry W, Behki
development of Reference Title: Rhodococcus rec of dealkylating and Reference Title: dechlorinating the Reference Author: Shao ZQ, Seffe	ens W, Mulbry W, Behki
development of Reference Title: Rhodococcus rec of dealkylating and Reference Title: dechlorinating th Reference Author: Shao ZQ, Seffe RM; Reference Location: J Bacteriol 198	ens W, Mulbry W, Behki 95;177:5748-5755.
development of Reference Title: Rhodococcus rec of dealkylating and Reference Title: dechlorinating the Reference Author: Shao ZQ, Seffe RM; Reference Location: J Bacteriol 198 Database Reference INTERPRO;	ens W, Mulbry W, Behki 95;177:5748-5755. IPR002604;
development of Reference Title: Rhodococcus rec of dealkylating and Reference Title: dechlorinating the Reference Author: Shao ZQ, Seffe RM; Reference Location: J Bacteriol 199 Database Reference INTERPRO; Database reference: PFAMB; PB03	ens W, Mulbry W, Behki 95;177:5748-5755. IPR002604; 84853;
development of Reference Title: Rhodococcus rec of dealkylating and Reference Title: dechlorinating the Reference Author: Shao ZQ, Seffe RM; Reference Location: J Bacteriol 198 Database Reference INTERPRO; Database reference: PFAMB; PB03 Database reference: PFAMB; PB03	ens W, Mulbry W, Behki 95;177:5748-5755. IPR002604; 84853;



		221
Pfam Prosite	Full Name	Description
		Comment: these enzymes catalyse hydrolytic dechlorination of their substrates.
		Comment: Atrazine chlorohydrolase (AtzA) from
		Pseudomonas sp. Swiss:P72156
		Comment: catalyses the dechlorination of atrazine to
		hydroxyatrazine [1]. Comment: s-Triazine hydrolase (TrzA) form R.
		corallinus Swiss:P72156
•		Comment: catalyses the deamination and dechlorination
		of melamine and
	,	Comment: deethylsimazine to ammeline and N- ethylammeline [1].
		Number of members: 29
B56	Protein phosphatase 2A	Accession number: PF01603
	regulatory B subunit (B56 family)	, , , , , , , , , , , , , , , , , , , ,
İ	Tairilly)	(B56 family) Author: Bateman A
		Alignment method of seed: Clustalw
		Source of seed members: Pfam-B_984 (release 4.1)
		Gathering cutoffs: 11 11
		Trusted cutoffs: 17.80 17.80 Noise cutoffs: 5.50 5.50
		HMM build command line: hmmbuild -f HMM SEED
		HMM build command line: hmmcalibrateseed 0 HMM
		Reference Number: [1] Reference Medline: 96064678
		Reference Medline: 96064678 Reference Title: Identification of a new family of protein
		phosphatase 2A
†		Reference Title: regulatory subunits.
		Reference Author: McCright B, Virshup DM;
		Reference Location: J Biol Chem 1995;270:26123-26128. Database Reference INTERPRO; IPR002554;
		Comment: Protein phosphatase 2A (PP2A) is a major
		intracellular protein
		Comment: phosphatase that regulates multiple aspects
		of cell growth and metabolism. Comment: The ability of this widely distributed
		heterotrimeric enzyme to act on a
		Comment: diverse array of substrates is largely
		controlled by the nature of its
		Comment: regulatory B subunit. There are multiple families of B subunits (See also
		Comment: PR55), this family is called the B56 family
		[1].
		Number of members: 34
Bac_export_1	Bacterial export proteins,	Accession number: PF01311
	family 1	Definition: Bacterial export proteins, family 1 Author: Finn RD, Bateman A
		Alignment method of seed: Clustalw
		Source of seed members: Pfam-B_1442 (release 3.0)
		Gathering cutoffs: 25 25
		Trusted cutoffs: 37.20 37.20 Noise cutoffs: -95.00 -95.00
		HMM build command line: hmmbuild -F HMM SEED
		HMM build command line: hmmcalibrateseed 0 HMM
		Reference Number: [1]
		Reference Medline: 95113771 Reference Title: Caulobacter FliQ and FliR membrane
		proteins, required for
		Reference Title: flagellar biogenesis and cell division, belong
		to a family
		Reference Title: of virulence factor export proteins. Reference Author: Zhuang WY, Shapiro L:
		Reference Author: Zhuang WY, Shapiro L; Reference Location: J Bacteriol 1995;177:343-356.
		Database Reference INTERPRO; IPR002010;
		Comment: This family includes the following members;
		Comment: FliR, MopE, SsaT, YopT, Hrp, HrcT and
		SpaR Comment: All of these members export proteins, that
		do not possess signal
		Comment: peptides, through the membrane. Although
. [the proteins that these



Pfam	Prosite	Full Name	Description
			Comment: exporters move may be different, the exporters are thought to Comment: function in similar ways [1].
			Number of members: 29
Band_41	PDOC00566	Band 4.1 family domain signatures and profile	A number of cytoskeletal-associated proteins that associate with various
			proteins at the interface between the plasma membrane and the cytoskeleton
			contain a conserved N-terminal domain of about 150 amino-acid residues [1,2,
			3]. The proteins in which such a domain is known to exist are listed below.
			- Band 4.1, which links the spectrin-actin cytoskeleton of erythrocytes to
			the plasma membrane. Band 4.1 binds with a high affinity to glycophorin and
			with lower affinity to band 3 protein Ezrin (cytovillin or p81), a component of the undercoat of the microvilli
			plasma membrane.
			- Moesin, which is probably involved in binding major cytoskeletal structures
			to the plasma membrane Radixin, which seems to play a crucial role in the binding of
			the barbed end of actin filaments to the plasma membrane in the undercoat
			of the cell- to-cell adherens junction (AJ).
			- Talin, which binds with high affinity to vinculin and with low affinity to
			integrins. Talin is a high molecular weight (270 Kd) cytoskeletal protein
			concentrated in regions of cell-substratum contact and, in lymphocytes, of
			cell-cell contacts Filopodin, a slime mold protein that binds actin ans which is
			the control of cell motility and chemotaxis.
			- Merlin (or schwannomin). Defects in this protein are the cause of type 2
			neurofibromatosis (NF2), a predisposition to tumors of the nervous system.
			- Protein NBL4.
			- Protein-tyrosine phosphatases PTPN3 (PTP-H1) and PTPN4 (PTP-MEG1).
			Structurally these two very similar enzymes are composed of a N-terminal
			band 4.1-like domain followed by a central segment of unknown function and
			a C-terminal catalytic domain (see <pdoc00323>). They could act at</pdoc00323>
			junctions between the membrane and the cytoskeleton Protein-tyrosine phosphatases PTPN14 (PEZ or PTP36) and
			PTP-D1, PTP-RL10 and PTP2E. These phosphatases also consist of a N-terminal
			band 4.1-like domain and a C-terminal catalytic domain. The central
			domain seems to contain a SH3-binding domain.
			- Caenorhabditis elegans protein phosphatase ptp-1.
			Ezrin, moesin, and radixin are highly related proteins, but the other proteins in which this domain is found do not share any region of similarity
			outside of the domain. In band 4.1 this domain is known to be
		Į.	important for the interaction with glycophorin, an integral membrane protein.
			We have developed two signature patterns for this domain, one is
			pased on the



			353
Pfam	Prosite	Full Name	Description
			conserved positions found at the N-terminal extremity of the domain, the second is located in the C-terminal section.
			Description of pattern(s) and/or profile(s) Consensus pattern W-[LIV]-x(3)-[KRQ]-x-[LIVM]-x(2)-[QH]-x(0,2)-
			[LIVMF]- x(6,8)-[LIVMF]-x(3,5)-F-[FY]-x(2)-[DENS] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern [HYW]-x(9)-[DENQSTV]-[SA]-x(3)-[FY]- [LIVM]-x(2)-[ACV]- x(2)-[LM]-x(2)-[FY]-G-x-[DENQST]-[LIVMFYS] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT 7.
			Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Expert(s) to contact by email Rees J. jrees@vax.oxford.ac.uk
			Last update November 1997 / Patterns and text revised; profile added. References [1]
			Rees D.J.G., Ades S.A., Singer S.J., Hynes R.O. Nature 347:685-689(1990).
			[2] Funayama N., Nagafuchi A., Sato N., Tsukita S., Tsukita S. J. Cell Biol. 115:1039-1048(1991).
			[3] Takeuchi K., Kawashima A., Nagafuchi A., Tsukita S. J. Cell Sci. 107:1921-1928(1994).
biotin_lipoyl	PDOC00167; PDOC00168	Biotin-requiring enzymes; 2-oxo acid dehydrogenases acyltransferase component lipoyl binding	Biotin, which plays a catalytic role in some carboxyl transfer reactions, is covalently attached, via an amide bond, to a lysine residue in enzymes requiring this coenzyme [1,2,3,4]. Such enzymes are:
			- Pyruvate carboxylase (EC 6.4.1.1) Acetyl-CoA carboxylase (EC 6.4.1.2) Propionyl-CoA carboxylase (EC 6.4.1.3) Methylcrotonoyl-CoA carboxylase (EC 6.4.1.4) Geranoyl-CoA carboxylase (EC 6.4.1.5) Urea carboxylase (EC 6.3.4.6) Oxaloacetate decarboxylase (EC 4.1.1.3) Methylmalonyl-CoA decarboxylase (EC 4.1.1.41) Glutaconyl-CoA decarboxylase (EC 4.1.1.70) Methylmalonyl-CoA carboxyl-transferase (EC 2.1.3.1) (transcarboxylase).
			Sequence data reveal that the region around the biocytin (biotin-lysine) residue is well conserved and can be used as a signature pattern.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [GN]-[DEQTR]-x-[LIVMFY]-x(2)-[LIVM]-x-[AIV]-M-K-[LMAT]- x(3)-[LIVM]-x-[SAV] [K is the biotin attachment site]



		C	354
Pfam	Prosite	Full Name	Description
			Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Note the domain around the biotin-binding lysine residue is evolutionary related to that around the lipoyl-binding lysine residue of 2-oxo acid dehydrogenase acyltransferases (see <pdoc00168>).</pdoc00168>
			Last update November 1997 / Pattern and text revised. References
			[1] Knowles J.R. Annu. Rev. Biochem. 58:195-221(1989).
			[2] Samols D., Thronton C.G., Murtif V.L., Kumar G.K., Haase F.C., Wood H.G. J. Biol. Chem. 263:6461-6464(1988).
			[3] Goss N.H., Wood H.G. Meth. Enzymol. 107:261-278(1984).
			[4] Shenoy B.C., Xie Y., Park V.L., Kumar G.K., Beegen H., Wood H.G., Samols D. J. Biol. Chem. 267:18407-18412(1992).
			The 2-oxo acid dehydrogenase multienzyme complexes [1,2] from bacterial and eukaryotic sources catalyze the oxidative decarboxylation of 2-oxo acids to the corresponding acyl-CoA. The three members of this family of multienzyme
			- Pyruvate dehydrogenase complex (PDC) 2-oxoglutarate dehydrogenase complex (OGDC) Branched-chain 2-oxo acid dehydrogenase complex (BCOADC).
			These three complexes share a common architecture: they are composed of multiple copies of three component enzymes - E1, E2 and E3. E1 is a thiamine pyrophosphate-dependent 2-oxo acid dehydrogenase, E2 a dihydrolipamide acyltransferase, and E3 an FAD-containing dihydrolipamide dehydrogenase.
			E2 acyltransferases have an essential cofactor, lipoic acid, which is covalently bound via a amide linkage to a lysine group. The E2 components of OGCD and BCOACD bind a single lipoyl group, while those of PDC bind either one (in yeast and in Bacillus), two (in mammals), or three (in Azotobacter and in Escherichia coli) lipoyl groups [3].
			In addition to the E2 components of the three enzymatic complexes described above, a lipoic acid cofactor is also found in the following proteins:
			- H-protein of the glycine cleavage system (GCS) [4]. GCS is a multienzyme complex of four protein components, which catalyzes the degradation of glycine. H protein shuttles the methylamine group of glycine from the P
			from the P protein to the T protein. H-protein from either prokaryotes or eukaryotes binds a single lipoic group.

			355
Pfam	Prosite	Full Name	Description - Mammalian and yeast pyruvate dehydrogenase complexes differ from that of other sources, in that they contain, in small amounts, a protein of unknown function - designated protein X or component X. Its sequence is closely related to that of E2 subunits and seems to bind a lipoic group [5]. - Fast migrating protein (FMP) (gene acoC) from Alcaligenes eutrophus [6].
			This protein is most probably a dihydrolipamide acyltransferase involved in acetoin metabolism. We developed a signature pattern which allows the detection of the lipoylbinding site.
			Description of pattern(s) and/or profile(s) Consensus pattern [GN]-x(2)-[LIVF]-x(5)-[LIVFC]-x(2)-[LIVFA]-x(3)-K-[STAIV]- [STAVQDN]-x(2)-[LIVMFS]-x(5)-[GCN]-x-[LIVMFY] [K is the lipoyl-binding site] Sequences known to belong to this class detected by the pattern ALL.
-			Other sequence(s) detected in SWISS-PROT 2. Note the domain around the lipoyl-binding lysine residue is evolutionary related to that around the biotin-binding lysine residue of biotin requiring enzymes (see <pdoc00167>). Last update November 1995 / Text revised. References [1] Yeaman S.J. Biochem. J. 257:625-632(1989).</pdoc00167>
			[2] Yeaman S.J. Trends Biochem. Sci. 11:293-296(1986). [3] Russel G.C., Guest J.R. Biochim. Biophys. Acta 1076:225-232(1991).
			Fujiwara K., Okamura-Ikeda K., Motokawa Y. J. Biol. Chem. 261:8836-8841(1986). [5] Behal R.H., Browning K.S., Hall T.B., Reed L.J. Proc. Natl. Acad. Sci. U.S.A. 86:8732-8736(1989). [6] Priefert H., Hein S., Krueger N., Zeh K., Schmidt B., Steinbuechel A.
Biotin_synth		Biotin synthase	J. Bacteriol. 173:4056-4071(1991). Accession number: PF01792 Definition: Biotin synthase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1407 (release 4.2) Gathering cutoffs: -180 -180 Trusted cutoffs: -180 -176.30 Noise cutoffs: -183.90 -183.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 96312354 Reference Title: Cloning, sequencing, and characterization of the Bacillus



Prosite Full Name Description Reference Title: subtilis biotin biosynthetic of Reference Author: Bower S, Perkins JB, You CL, Rahaim P, Pero J; Reference Location: J Bacteriol 1996;178:412 Reference Number: [2] Reference Mediline: 97074643 Reference Title: Two new members of the bioloning, Reference Title: sequencing and expression of Methylobacillus Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. INTERPRO; IPR002684 Database reference: Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, comment: and possibly cysteine to contobiotin [1]. Comment: Biotin (vitamin H) is a prosth	cum RR, Howitt 22-4130. io B superfamily: a of bio B genes terium f, Tsygankov k;
Reference Author: CL, Rahaim P, Pero J; Reference Location: Reference Number: Reference Number: Reference Medline: Reference Title: Cloning, Reference Title: Reference Title: Sequencing and expression of Methylobacillus Reference Title: Glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Database Reference Database Reference: Database reference: Database	cum RR, Howitt 22-4130. In of bio B genes Iterium If, Tsygankov It;
CL, Rahaim P, Pero J; Reference Location: J Bacteriol 1996;178:412 Reference Number: [2] Reference Medline: 97074643 Reference Title: Two new members of the bicloning, Reference Title: sequencing and expression of Methylobacillus Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference: INTERPRO; IPR002684 Database reference: PFAMB; PB040740; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to conto biotin [1].	22-4130. io B superfamily: n of bio B genes terium f, Tsygankov i; works with
Reference Location: Reference Number: [2] Reference Medline: 97074643 Reference Title: Two new members of the bicloning, Reference Title: sequencing and expression of Methylobacillus Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference: INTERPRO; IPR002684 Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to conto biotin [1].	io B superfamily: n of bio B genes terium f, Tsygankov l; works with
Reference Number: [2] Reference Medline: 97074643 Reference Title: Two new members of the bicloning, Reference Title: sequencing and expression of Methylobacillus Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference: INTERPRO; IPR002684 Database reference: PFAMB; PB043954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to conto biotin [1].	io B superfamily: n of bio B genes terium f, Tsygankov l; works with
Reference Medline: 97074643 Reference Title: Two new members of the bicloning, Reference Title: sequencing and expression of Methylobacillus Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference: INTERPRO; IPR002684 Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 vflavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to conto biotin [1].	n of bio B genes terium 1, Tsygankov 1; works with
Reference Title: Two new members of the bicloning, Reference Title: sequencing and expression of Methylobacillus Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference INTERPRO; IPR002684 Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to conto biotin [1].	n of bio B genes terium 1, Tsygankov 1; works with
cloning, Reference Title: sequencing and expression of Methylobacillus Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference INTERPRO; IPR002684 Database reference: PFAMB; PB040740; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to con to biotin [1].	n of bio B genes terium 1, Tsygankov 1; works with
Reference Title: sequencing and expression of Methylobacillus Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference: INTERPRO; IPR002684 Database reference: PFAMB; PB040740; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to conto biotin [1].	terium I, Tsygankov I; works with
of Methylobacillus Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference INTERPRO; IPR002684 Database reference: PFAMB; PB023954; Database reference: PFAMB; PB041740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to con to biotin [1].	terium I, Tsygankov I; works with
Reference Title: flagellatum and Corynebact glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference: INTERPRO; IPR002684 Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to conto biotin [1].	f, Tsygankov I; works with
glutamicum. Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference INTERPRO; IPR002684 Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to conto biotin [1].	f, Tsygankov I; works with
Reference Author: Serebriiskii IG, Vassin VM YD; Reference Location: Gene 1996;175:15-22. Database Reference INTERPRO; IPR002684 Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to con to biotin [1].	k; works with
YD; Reference Location: Gene 1996;175:15-22. Database Reference INTERPRO; IPR002684 Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to conto biotin [1].	k; works with
Reference Location: Gene 1996;175:15-22. Database Reference INTERPRO; IPR002684 Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to con to biotin [1].	works with
Database Reference INTERPRO; IPR002684 Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to con to biotin [1].	works with
Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to con to biotin [1].	works with
Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to con to biotin [1].	
Comment: Biotin synthase EC:2.8.1.6 v flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to con to biotin [1].	
flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to con to biotin [1].	
Comment: and possibly cysteine to con to biotin [1].	
to biotin [1].	and the second s
	ivert detniobiotin
Comment: Biotin (vitamin H) is a prosth	
	etic group in
enzymes catalysing	
Comment: carboxylation and transcarbo	oxylation
reactions [2].	
Number of members: 29	
BolA BolA-like protein Accession number: PF01722	
BolA BolA-like protein Accession number: PF01722 Definition: BolA-like protein	
Author: Bashton M, Bateman A	
Alignment method of seed: Clustalw	
Source of seed members: Pfam-B 1996 (release	a 4 1)
Gathering cutoffs: 23 23	,
Trusted cutoffs: 23.70 23.70	
Noise cutoffs: -16.00 -16.00	
HMM build command line: hmmbuild -F HMM SE	ED
HMM build command line: hmmcalibrateseed 0) HMM
Reference Number: [1]	
Reference Medline: 99291046	
Reference Title: The stationary-phase morph	nogene bolA
from Escherichia coli	
Reference Title: is induced by stress during of	early stages of
growth.	-4- 44
Reference Author: Santos JM, Freire P, Vicer	nte M, Arraiano
	200
Reference Location: Mol Microbiol 1999;32:78 Reference Number: [2]	J-130.
Reference Medline: 90059998	
Reference Title: Induction of a growth-phase	a-dependent
promoter triggers	. soporiuonit
Reference Title: transcription of bolA, an Esc	cherichia coli
morphogene.	
Reference Author: Aldea M, Garrido T, Herna	andez-Chico C.
Vicente M, Kushner	,
Reference Author: SR;	
Reference Location: EMBO J 1989;8:3923-393	
Database ReferenceINTERPRO; IPR002634	•
Comment: This family consist of the mo	orpho-protein
BolA from	
Comment: E. coli and its various homol	ogs. In E. coli
over expression of	orobolog:
Comment: this protein causes round mo	orphology and
may be involved in	donastion and
Comment: switching the cell between e	iongation and
septation systems during Comment: cell division [1]. The express	sion of BotA is
growth rate regulated	אוטם וט ווטוא ו
Comment: and is induced during the tra	ansition into the
the stationary	
Comment: phase [1]. BolA is also induc	ed by stress
during early stages of	-,
Comment: growth [1] and may have a g	

			357
Pfam.	Prosite	Full Name	Description stress response. Comment: It has also been suggested that BolA can induce the transcription Comment: of penicillin binding proteins 6 and 5 [2,1]. Number of members: 18
casein_kappa			Accession number: PF00997 Definition: Kappa casein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1298 (release 3.0) Gathering cutoffs: -32 -32 Trusted cutoffs: 16.40 16.40 Noise cutoffs: -73.00 -73.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: [1] Reference Medline: 98072500 Reference Title: evidence for positive selection within the family Bovidae. Reference Author: Ward TJ, Honeycutt RL, Derr JN; Reference Location: Database Reference Comment: Genetics 1997;147:1863-1872. INTERPRO; IPR000117; Kappa-casein is a mammalian milk protein involved in a Comment: number of important physiological processes. In the gut, Comment: (para kappa-casein) and a soluble hydrophilic glycopeptide Comment: (caseinomacropeptide). Caseinomacropeptide is responsible Comment: for increased efficiency of digestion, prevention of neonate Comment: hypersensitivity to ingested proteins, and inhibition of Comment: gastric pathogens. Number of members: 56
CAT	PDOC00093	Chloramphenicol acetyltransferase	Chloramphenicol acetyltransferase (CAT) (EC 2.3.1.28) [1] catalyzes the acetyl-CoA dependent acetylation of chloramphenicol (Cm), an antibiotic which inhibits prokaryotic peptidyltransferase activity. Acetylation of Cm by CAT inactivates the antibiotic. A histidine residue, located in the C-terminal section of the enzyme, plays a central role in its catalytic mechanism. We derived a signature pattern from the region surrounding this active site residue. Description of pattern(s) and/or profile(s) Consensus pattern Q-[LIV]-H-H-[SA]-x(2)-D-G-[FY]-H [The second H is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note there is a second family of CAT [2], evolutionary unrelated to the main family described above. These CAT belong to the bacterial hexapeptide-repeat containing-transferases family (see <pdoc00094>). Last update November 1997 / Text revised. References [1]</pdoc00094>



Shaw W.V., Leslie A.G.W. Annu. Rev. Biophys. Chem. 20:363-386(1991). [2] Parent R., Roy P.H. J. Bacteriol. 174:2891-2897(1992). Cation_efflux Cation efflux family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_232 (release 4.0) Gathering outoffs: 6-90 e.90 Noise cutoffs: 1-9.30 -19.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Mediline: [98361887] Reference Title: conferring resistance to zinc and cobalt ions in Reference Title: Staphylococcus aureus. Reference Author: Xiong A, Jayaswal RK; Reference Author: Xiong A, Jayaswal RK; Reference Number: [2] Reference Mediline: Conferring resistance to zinc and cobalt ions in Reference Title: Staphylococcus aureus. Reference Author: Xiong A, Jayaswal RK; Reference Author: Sing A, Jayaswal RK; Reference Number: [2] Reference Mediline: [30, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,				0.00
Annu. Rev. Biophys. Chem. 20:383-386(1991). 2	Pfam	Prosite	Full Name	Description
Cation_efflux Cation_efflux Cation_efflux family Accession number: PF01545 Definition: Cation efflux family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfami B_ 232 (release 4.0) Gambering and 33: Gambering and 33: Gambering and 33: Gambering and 33: Gambering and 33: Gambering and 33: Gambering and 33: Gambering and 33: Gambering and 33: Gambering and 34: Gamber				
Cation_efflux Cation				Annu. Rev. Biophys. Chem. 20:363-386(1991).
Cation_efflux Cation				(0)
Cation efflux family Accession number: PF01545 Cation efflux family Accession number: PF01545 Cation efflux family Alignment method of seed: Clustatu Source of seed members: Pfam-B 232 (release 4.0) Gathering cutoffs: -6 -6 Trusted cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Noise cutoffs: -19,30 -19,30 Reference Number: Reference Author: Reference Auth			i	
Cation efflux family Accession number: PF01545 Definition: Cation efflux family Author: Bateman A Alignment method of seed: Clustain Gathering cutoffs: 6.96.89 Noise cutoffs: 19.30-19.30 HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-F HMM SEED HMM build command line: hmmbuild-HMM SEED HMM build command li				
Definition: Cation efflux family Author: Bateman A Alignment method of seed: Clustatus Source of seed members: Pfamile, 232 (release 4.0) Gathering outoffs: 6.9 6.9 7 Instead outoffs: 6.90 6.90 New South March Command Inc. Instead Instead of the March Command Inc. Instead Inste				J. Dacterior. 174.2091-2097(1992).
Definition: Cation efflux family Author: Bateman A Alignment method of seed: Clustatus Source of seed members: Pfamile, 232 (release 4.0) Gathering outoffs: 6.9 6.9 7 Instead outoffs: 6.90 6.90 New South March Command Inc. Instead Instead of the March Command Inc. Instead Inste	Cation efflux		Cation efflux family	Accession number: PF01545
Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Plam-B_232 (release 4.0) Gathering culoffs: 6-86 Trusted cutoffs: 6-96 6	Oddon_chiax		Cation chiax family	
Alignment method of seed: Clustalw Source of seed members: Pfram—B_232 (release 4.0) Gathering cutoffs: 6-96 e.9 Trusted cutoffs: 6-96 e.9 Noise cutof				I
Source of seed members: Pfam-B_232 (release 4.0) Gathering cutoffs: 6-6 Trusted outoffs: 6-90-690 Noise cutoffs: 19,30-19,30 HMM build command line: hummballbrate-seed 0 HMM Programment of the members: Pfame Programment Pr		:		
Gathering cutoffs: 6-96 90 Noise cutoffs: 1-93 0-19-30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM Seference Mediline: 98381887 Reference Mediline: Reference Title: confering resistance to zinc and cobalt ions Staphylococcus aureus. Reference Author: Xing A, Jayaswal RK; Reference Author: Seference Mediline: Reference Mediline: Reference Author: S				
Trusted outoffs: 19.30 :19.30 :19.30 Noise cutoffs: 19.30 :19.30 :19.30 HMM build command line: hmmbuild -F.HMM SEED HMM build command line: hmmbuild -F.HMM SEE				1.
Noise cutoffs: 1-9,30-19,30 HMM build command line: hmmbuild F-HMM SEED HMM build command line: hmmbuild F-HMM SEED HMM build command line: hmmbuild F-HMM SEED HMM build command line: hmmbuild F-HMM SEED Reference Mediline: Reference Mediline: Reference Mediline: Reference Mediline: Reference Title: conferring resistance to zinc and cobalt ions Reference Title: Staphylococcus aureus Reference Title: JBacteriol 1998;180-4024-4029. Reference Mumber: Reference Mediline: Reference Mediline: Reference Mediline: Reference Mediline: Reference Author: Kanazawa S, Reference Author: Kanazawa S, Reference Author: Kanazawa S, Reference Author: Kanazawa S, Reference Author: Reference Author: Kanazawa S, Reference Author: Reference Reference Procession Procession Reference Refe				
HMM build command line: hmmbuild r. HMM SEED HMM build command line: hmmbuild rate -seed of MIM Reference Number: [1] Reference Mumber: [1] Reference Mumber: [1] Molecular characterization of a chromosomal determinant Reference Title: line: Reference Title: Reference Author: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Title: genes in Alcaligenes Reference Title: Reference Author: Kanazawa S, Reference Location: 704. Database Reference: Comment: Intellection of the Seed of the Se				
HMM build command line: hmmcalibrateseed 0 HMM Reference Number: Reference Mediline: 98361887 Reference Title: shromosomal determinant Reference Title: normosomal determinant Reference Title: Reference Author: Reference Author: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Mumber: Reference Author: Kanazawa S, Reference Author: Reference Author: Reference Author: Reference Author: Reference Location: 704. Database Reference: Database reference: Database reference: Comment: membrane proteins, that Comment: are found to increase tolerance to divalent metal ions such Comment: be efflux pumps that remove these ions from cells. Number of members: 59 CBD_6 Cellulose binding domain Author: Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19. 10. 19. 10. 10. 10. 10. 10. 10. 10. 10. 10. 10				
Reference Multimer Reference Title: normand determinant Reference Title: normand determinant Reference Title: normand determinant Reference Author: Reference Author: Reference Author: Reference Author: Reference Medine: Reference Medine: Reference Medine: Reference Medine: Reference Multimer: Reference Medine: Reference Author: Reference Database Reference: Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: membrane proteins, that Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: sa cadmium, zinc, and cobalt. These proteins are thought to Comment: Sa cadmium, zinc, and cobalt. These proteins are thought to Comment: Sa cadmium, zinc, and cobalt. These proteins are thought to Comment: Sa cadmium, zinc, and cobalt. These proteins, that are found to increase tolerance to division and the proteins, that are found to increase tolerance to division and the proteins, that are found to increase tolerance to division and the proteins, that are found to increase tolerance to division and the proteins, that are found to increase tolerance to division and the proteins, that are found to increase tolerance to division				
Reference Mediline: 98361887 Reference Title: northornosomal determinant Reference Title: northornosomal determinant Reference Title: northornosomal determinant Reference Title: northornosomal determinant Reference Title: northornosomal determinant Reference Title: northornosomal determinant Reference Title: northornosomal determinant Reference Title: northornosomal determinant Reference Author: Reference Number: pace and Northornosomal Reference Title: genes in Alcaligenes and Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Detabase Reference: Database Reference: Database Reference Database reference: Comment: proteins are thought to Comment in the proteins are thought to Comment in the proteins are thought to Comment in the proteins are thought to Commend in Comment in the proteins are thought to Comment in the proteins are thought to Commend in the proteins are thought to Commend in the proteins are thought to Commend in the proteins are thought to Commend in the proteins are thought to Commend in the proteins are thought to Commend in the proteins are thought to Commend in the proteins are thought to Commend in the proteins are thought to Commend in the proteins are thought to Commend in the proteins are thou				
Reference Title: chromosomal determinant Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Medine: Reference Author: Reference Medine: Reference Medine: Reference Medine: Reference Muthor: Reference Muthor: Reference Muthor: Reference Author: Reference Mumber: Referenc				1
chromosomal determinant Reference Title: Reference Title: Reference Location: Reference Mediline: Reference Mediline: Reference Mediline: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Database Reference Database Reference Database reference: Database Reference Title: Reference Ritle: Reference				
Reference Title: Staphylcoccus aureus. Stophylcochus apselbed Stophylcochus apselbed Stophylcoccus aureus. Stophylcochus apselbed Stophylcochus apselbed Stophylcochus apselbed Stophylcochus apselbed Stophylcochus apselbed Stophylcochus apselbed Stophylcochus apselbed Stophylcochus apse				
In Reference Title: Reference Location: Reference Decation: Reference Meditine: Reference Meditine: Reference Meditine: Reference Meditine: Reference Meditine: Reference Author: Kanazawa S, Reference Author: Kanazawa S, Reference Database Reference Database reference: DDB; fullo ; 1; 149;				
Reference Author: Reference Author: Reference Number: Reference Number: Reference Number: Reference Number: Reference Middine: Reference Middine: Reference Middine: Reference Middine: Reference Middine: Reference Middine: Reference Title: Reference Author: Reference Rediline: Reference Rediline: Reference Rediline: Reference Rediline: Reference Rediline: Reference Rediline: Reference Rumber: Reference Rum				l.
Reference Author: Reference Location: Reference Number: Reference Medline: Reference Title: Reference Title: Reference Title: Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Location: 704 Database Reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database Reference Database Reference Database Reference Database reference: Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference PDBSJMM] Database Reference PDBSJMM] Database Reference Database Reference PDBSJMM] Database Reference PDBSJMM] Database Reference PDBS, Tulo; 1; 149;				
Reference Number: Reference Number: Reference Number: Reference Mediline: Reference Mediline: Reference Title: genes in Alcaligenes Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Location: 704. Database Reference: Database Reference: Database Reference: Database Reference: Comment: membrane proteins, that Comment: metal ions such Comment: ac acamium, zinc, and cobalt. These proteins are thought to Comment: be tefflux pumps that remove these ions from cells. Number of members: Special Support of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19 10 19 10 Noise cutoffs: Reference Medline: Reference Medline: Reference Title: Reference				
Reference Number: Reference Medinie: genes in Alcaiglenes Reference Title: genes in Alcaiglenes Reference Author: Kanazawa S, Reference Author: Reference Author: Reference Author: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Reference Database reference: Database reference: Reference Number: Reference Reference Location: Reference Number: Reference Number: Reference Number: Reference Number: Reference Title: Ref				
Reference Medline: Reference Title: Reference Title: Reference Title: Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Possion Reference: Database Reference: Database Reference: Reference Possion Reference: Reference Reference: Reference Reference: Reference Reference: Reference Reference: Reference Reference: Reference Reference: Reference Reference: Reference Title: Reference Ruthor: Refe			1	
Reference Title: genes in Alcaligenes Reference Title: genes in Alcaligenes Reference Patters Reference Author: Kanazawa S, Reference Author: Reference Author: Reference Location: 704. Database Reference Database Reference: Database Reference: Comment: membrane proteins, that Comment: are found to increase tolerance to divalent metal ions such Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: Selfulose binding domain Author: Alignment method of seed. Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19				
Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Location: 704. Database Reference Database Reference Database reference: Comment: membrane proteins, that Comment: are found to increase tolerance to divalent metal ions such Comment: be efflux pumps that remove these ions from cells. Number of members: CBD_6 Cellulose binding domain Author: Cellulose binding domain Author:				Reference Title: Cloning and sequence analysis of czc
Reference Author: Kanazawa S, Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Location: 704. Database Reference: Database reference: Database reference: Comment: Members of this family are integral membrane proteins, that Comment: metal ions such Comment: proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: CBD_6 Cellulose binding domain Accession number: Definition: Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19 10 19 10 Noise cutoffs: 8,90 8,90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbaild reference Number: Reference Reference Title: domain of Reference Title: ruclear magnetic Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Ruthor: DG, McIntosh LP; Reference Location: DG, McIntosh LP; Reference Locati				genes in Alcaligenes
Kanazawa S, Reference Author: Reference Location: 704. Database Reference Database Reference Database reference: Comment: membrane proteins, that Comment: metal ions such Comment: proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: CBD_6 Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19 10 19-10 Noise cutoffs: 8-90 8-90 HMM build command line: hmmbuild HMM SEED HMM build				1 ²
Reference Author: Reference Location: Potabase Reference Database Reference: Database reference: Database reference: Comment: Members of this family are integral membrane proteins, that Comment: metal ions such Comment: proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: CBD_6 Cellulose binding domain Accession number: Definition: Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19 10 19.10 19.10 Noise cutoffs: 19 0 Noise cutoffs: 19 0 Noise cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: domain of Reference Title: nuclear magnetic Reference Title: Definition: Definition: Reference Coulon: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Location: Definition: Database Reference: http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM				Reference Author: Kunito T, Kusano T, Oyaizu H, Senoo K,
Reference Author: Reference Location: Potabase Reference Database Reference: Database reference: Database reference: Comment: Members of this family are integral membrane proteins, that Comment: metal ions such Comment: proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: CBD_6 Cellulose binding domain Accession number: Definition: Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19 10 19.10 19.10 Noise cutoffs: 19 0 Noise cutoffs: 19 0 Noise cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 0 Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: domain of Reference Title: nuclear magnetic Reference Title: Definition: Definition: Reference Coulon: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Location: Definition: Database Reference: http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM Database Reference: PDBSUM				Kanazawa S,
TO4. Database Reference Database reference: Database reference: Database reference: Database reference: Database reference: Defence of this family are integral membrane proteins, that Comment: metal ions such Comment: metal ions such Comment: proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: Definition: Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Definition: Source of seed members: Definition: Database Reference: Noise cutoffs: Definition: De				
Database Reference INTERPRO; IPR002524; Database reference: PFAMB; PB038216; Comment: Members of this family are integral membrane proteins, that Comment: are found to increase tolerance to divalent metal ions such Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: 59 CBD_6 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19.0 Trusted cutoffs: 19.10 Noise cutoffs: 90 Trusted cutoffs: 90 HMM build command line: Inmbuild HMM SEED HMM build command line: Inmbuild HMM SEED HMM build command line: Inmbuild HMM SEED HMM Beference Number: [1] Reference Number: [1] Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Tresonance spectroscopy. Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Database Reference: URL; http://www.ocms.ox.ac.uk/-ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDBSUM] Database Reference PDB; 1ulo; 1; 149;				Reference Location: Biosci Biotechnol Biochem 1996;60:699-
Database reference: PFAMB; PB038216; Comment: Members of this family are integral membrane proteins, that Comment: are found to increase tolerance to divalent metal ions such Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: 59 CBD_6 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19.0 Trusted cutoffs: 19.10 19.10 Noise cutoffs: 19.10 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild reseed 0 HMM Reference Number: [1] Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Database Reference: URL; http://www.ocms.ox.ac.uk/-ponting/methmb/example.html; Database Reference: SCOP; Tulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDBSUM] Database Reference: PDB; Tulo; 1; 1,149;				704.
Comment: Members of this family are integral membrane proteins, that Comment: are found to increase tolerance to divalent metal ions such Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: 59 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.0 Trusted cutoffs: 8.90 8.90 HMM build command line: hambuild HMM SEED HMM build command line: hambuild HMM SEED HMM build command line: hambuild rateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: Tesonance spectroscopy. Reference Author: DG, McIntosh LP; Reference Location: Database Reference: http://www.ooms.ox.ac.uk/-ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDBS 1ulp; 1; 149;				Database Reference INTERPRO; IPR002524;
membrane proteins, that Comment: are found to increase tolerance to divalent metal ions such Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: 59 CBD_6 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19 10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM Reference Number: [1] Reference Number: [1] Reference Title: Office of the N-terminal cellulose-binding domain of Reference Title: Tellulomonas firmi CenC determined by nuclear magnetic Reference Title: Reference Title: Performed Location: Database Reference: Database Reference: Database Reference: Database Reference: Scop; Johnson PE, Joshi MD, Tomme P, Kilburn Database Reference: Scop; Julp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDBSUM] Database Reference: PDBS Julp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDBS Indo; f1; f149;				
Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: 59 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.10 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: domain of Reference Title: nuclear magnetic Reference Title: Reference Title: resonance spectroscopy. Johnson PE, Joshi MD, Tomme P, Kilburn Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				
Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: 59 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.10 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: domain of Reference Title: nuclear magnetic Reference Title: Reference Title: resonance spectroscopy. Johnson PE, Joshi MD, Tomme P, Kilburn Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				membrane proteins, that
Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: 59 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Structure of the N-terminal cellulose-binding domain of Reference Title: nuclear magnetic Reference Author: DG, McIntosh LP; Reference Author: DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. URL; Biochemistry 1996;35:14381-14394. URL; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				
CBD_6 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Beteman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19:0 Trusted cutoffs: 19:0 19:0 Noise cutoffs: 8:90 8:90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: Reference Title: Reference Title: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/-ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				metal ions such
Comment: be efflux pumps that remove these ions from cells. Number of members: 59 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19.0 Trusted cutoffs: 19.0 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: Immbuild HMM SEED HMM build command line: Immbuild HMM SEED HMM build command line: Immbuild HMM SEED HMM build command line: Structure of the N-terminal cellulose-binding domain of Reference Medline: 97074498 Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: Tesonance spectroscopy. Def, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;		:		Comment: as cadmium, zinc, and cobalt. These
CBD_6 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19 0 Trusted cutoffs: 19.0 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: Reference Title: Reference Sepectroscopy. Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo; 1; 149;				proteins are thought to
CBD_6 Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.0 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild + mmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: VRL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDSSUM] Database Reference: PDBS; 1ulo; 1; 1, 149;			İ	
Cellulose binding domain Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19 0 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: PDB; 1ulo; 1; 149;				cells.
Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.0 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo; 1; 149;				Number of members: 59
Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.0 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo; 1; 149;				
Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.10 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/-ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;	CBD 6		Cellulose binding domain	Accession number: PF02018
Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.10 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate seed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: PDBSUM] Database Reference: PDBSUM]	_			
Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				Author: Bateman A
Gathering cutoffs: 19 0 Trusted cutoffs: 19.10 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: DG, McIntosh LP; Reference Location: DG, McIntosh LP; Reference Location: Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				Alignment method of seed: Manual
Trusted cutoffs: 19.10 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				
Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				Gathering cutoffs: 19 0
HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;			1	
HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				
Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				
Reference Medline: Reference Title: Reference Title: Odmain of Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Author: DG, McIntosh LP; Reference Location: Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: PDBSUM] Database Reference: PDB; 1ulo; 1; 149;			l	
Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;		1		
domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo; 1; 149;				
Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				
nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;		1		
Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;				= 1
Reference Author: Johnson PE, Joshi MĎ, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo; 1; 149;				
DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo; 1; 149;	1			, , , ,
Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo; 1; 149;			1	1
Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo; 1; 149;		1	1	
http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo; 1; 149;			1	
Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo; 1; 149;				
PDBSUM] Database Reference PDB; 1ulo ; 1; 149;				
Database Reference PDB; 1ulo ; 1; 149;				
Database Reference PDB: 1ulp : 1: 149:				
1		1		Database Reference PDB; 1ulp ; 1; 149;



			859
Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1cx1 A; 2; 6;
			Database Reference PDB; 1ulo ; 150; 152;
			Database Reference PDB; 1ulp ; 150; 152;
	İ		Database Reference PDB; 1cx1 A; 7; 151;
			Database reference: PFAMB; PB012497;
			Database reference: PFAMB; PB041237;
			Database reference: PFAMB; PB041605;
			Number of members: 76
CBFD NFYB HMF	PDOC00578	CBF/NF-Y subunits	Diverse DNA hinding proteins are lineary to hind the COAAT
		signatures	Diverse DNA binding proteins are known to bind the CCAAT box, a common cis-
			acting element found in the promoter and enhancer regions of a
			large number of
			genes in eukaryotes. Amongst these proteins is one known as
			the CCAAT-binding
			factor (CBF) or NF-Y [1]. CBF is a heteromeric transcription
			factor that
		·	consists of two different components both needed for DNA-
			binding.
	,		The HAP protein complete of secret bit to the
			The HAP protein complex of yeast binds to the upstream activation site of
			cytochrome C iso-1 gene (CYC1) as well as other genes
i			involved in
			mitochondrial electron transport and activates their expression
			It also
			recognizes the sequence CCAAT and is structurally and
			evolutionary related to
			CBF.
			The first subunit of CBF, known as CBF-A or NF-YB in
			vertebrates, HAP3 in
			budding yeast and as php3 in fission yeast, is a protein of 116 to
			210 amino-
			acid residues which contains a highly conserved central domain
			of about 90
			residues. This domain seems to be involved in DNA-binding; we
			have developed a
			signature pattern from its central part.
			The second subunit of CBF, known as CBF-B or NF-YA in
			vertebrates, HAP2 in
			budding yeast and php2 in fission yeast, is a protein of 265 to 350
	i		amino-acid
			residues which contains a highly conserved region of about 60
	[residues. This
ļ			region, called the 'essential core' [2], seems to consist of two subdomains:
			an N-terminal subunit-association domain and a C-terminal
į	İ		DNA recognition
			domain. We have developed a signature pattern from a section
			of the subunit-
			association domain.
	· ·		Description of pattern(s) and/or profile(s)
			Consensus pattern C-V-S-E-x-I-S-F-[LIVM]-T-[SG]-E-A-[SC]-[DE]-[KRQ]-C
			Sequences known to belong to this class detected by the pattern
			ALL CBF-A subunits.
			Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern Y-V-N-A-K-Q-Y-x-R-I-L-K-R-R-x-A-R-A-K-L-E
1			Sequences known to belong to this class detected by the pattern
			ALL CBF-B subunits.
			Other sequence(s) detected in SWISS-PROT NONE.
1			Last update November 1995 / Patterns and text revised.
			rioronider 1333 / Fallents and text revised.
			References
			References
			References



Pfam Pi	rosite	Full Name	Description Nucleic Acids Res. 20:1087-1091(1992).
CbiX			Nucleic Acids Res. 20:1087-1091(1992).
CbiX			• •
CbiX			[2] Olesen J.T., Fikes J.D., Guarente L. Mol. Cell. Biol. 11:611-619(1991).
		CbiX	Accession number: PF01903
		CDIX	Definition: CbiX Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -25 -25 Trusted cutoffs: -23.10 -23.10 Noise cutoffs: -35.10 -35.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 98416126 Reference Title: Cobalamin (vitamin B12) biosynthesis: identification and Reference Title: characterization of a Bacillus megaterium cobl operon. Reference Author: Raux E, Lanois A, Warren MJ, Rambach A, Thermes C; Reference Location: Biochem J 1998;335:159-166.
			Database Reference Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: Database reference: DFAMB; PB040610; DFAMB; PB041575; The function of CbiX is uncertain, however it is found Comment: In cobalamin biosynthesis operons and so may have a Comment: related function. Some CbiX proteins
			contain a striking Comment: histidine-rich region at their C-terminus, which suggests
			Comment: that it might be involved in metal chelation [1].
United	DOCOGEGE	Changed budgeloops	Number of members: 6
cellulase Pí	DOC00565	Glycosyl hydrolases family 5 signature	The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family A [3] or as the glycosyl hydrolases family 5 [4,E1]. The enzymes which are currently known to belong to this family are listed below. - Endoglucanases from various species and strains of Bacillus Butyrivibrio fibrisolvens endoglucanases 1 (end1) and A (celA) Caldocellum saccharolyticum bifunctional endoglucanase/exoglucanase (celB). This protein consists of two domains; it is the C-terminal domain, which has endoglucanase activity, which belongs to this family Clostridium acetobutylicum endoglucanase (eglA) Clostridium cellulolyticum endoglucanases A (celccA) and D (celccD) Clostridium cellulovorans endoglucanase B (engB) and D (engD) Clostridium thermocellum endoglucanases B (celB), C (celC), E (celE), G (celG) and H (celH) Erwinia chrysanthemi endoglucanase Z (celZ) Fibrobacter succinogenes endoglucanase 3 (cel-3).



-			861
Pfam	Prosite	Full Name	Description
			 Pseudomonas solanacearum endoglucanase (egl). Robillarda strain Y-20 endoglucanase I. Ruminococcus albus endoglucanases I (EG-I), A (celA), and B (celB). Ruminococcus flavefaciens cellodextrinase A (celA). Ruminococcus flavefaciens endoglucanase E (celE). Streptomyces lividans endoglucanase. Thermomonospora fusca endoglucanase E-5 (celE). Trichoderma reesei endoglucanase II (EGLII). Xanthomonas campestris endoglucanase (engxcA). As well as:
		č	- Baker's yeast glucan 1,3-beta-glucosidase I/II (EC 3.2.1.58) (EXG1) Baker's yeast glucan 1,3-beta-glucosidase 2 (EC 3.2.1.58) (EXG2) Baker's yeast sporulation-specific glucan 1,3-beta-glucosidase (SPR1) Caldocellum saccharolyticum beta-mannanase (EC 3.2.1.78) (manA) Yeast hypothetical protein YBR056w Yeast hypothetical protein YIR007w. One of the conserved regions in these enzymes contains a conserved glutamic acid residue which is potentially involved [5] in the catalytic
			mechanism. We use this region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-x-N-E-[PV]- [RHDNSTLIVFY] [E is a putative active site residue] Sequences known to belong to this class detected by the pattern ALL, except for Robillarda Y-20 endoglucanase I whose sequence is known to be incorrect and yeast YBR056w. Other sequence(s) detected in SWISS-PROT 22. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr
			Last update November 1997 / Pattern and text revised. References [1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990). [2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).
			[3] Henrissat B., Claeyssens M., Tomme P., Lemesle L., Mornon JP. Gene 81:83-95(1989). [4] Henrissat B. Biochem. J. 280:309-316(1991).
н	PDOC00019	Actinin-type actin-binding	[5] Py B., Bortoli-German I., Haiech J., Chippaux M., Barras F. Protein Eng. 4:325-333(1991). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt
	. 5000019	domain signatures	Alpha-actinin is a F-actin cross-linking protein which is thought to anchor actin to a variety of intracellular structures [1]. The actin-binding

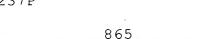


		,	362
Pfam	Prosite	Full Name	Description
			domain of alpha-actinin seems to reside in the first 250 residues of the
			protein. A
			similar actin-binding domain has been found in the N-terminal region of many
			different actin-binding proteins [2,3]:
			- In the beta chain of spectrin (or fodrin) In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD)
			and which may play a role in anchoring the cytoskeleton to the plasma
			membrane In the slime mold gelation factor (or ABP-120) In actin-binding protein ABP-280 (or filamin), a protein that link
			actin filaments to membrane glycoproteins In fimbrin (or plastin), an actin-bundling protein. Fimbrin differs
			from the above proteins in that it contains two tandem copies of the actin-
			binding domain and that these copies are located in the C- terminal part of the protein.
			We selected two conserved regions as signature patterns for this type of
			domain. The first of this region is located at the beginning of the domain, while the second one is located in the central section and has
			been shown to be essential for the binding of actin.
:			Description of pattern(s) and/or profile(s)
			Consensus pattern [EQ]-x(2)-[ATV]-[FY]-x(2)-W-x-N Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT 25.
			Consensus pattern [LIVM]-x-[SGN]-[LIVM]-[DAGHE]-[SAG]-x- [DNEAG]-[LIVM]-x- [DEAG]-x(4)-[LIVM]-x-[LM]-[SAG]-[LIVM]- [LIVMT]-W-x- [LIVM](2)
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update November 1997 / Patterns and text revised.
			References [1] Schleicher M., Andre E., Harmann A., Noegel A.A. Dev. Genet. 9:521-530(1988).
			[2] Matsudaira P. Trends Biochem. Sci. 16:87-92(1991).
			[3] Dubreuil R.R. BioEssays 13:219-226(1991).
chitinase_2	PDOC00839	Chitinases family 18 active site	Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence similarity chitinases belong to either family 18 or 19 in
			the classification of glycosyl hydrolases [2,E1]. Chitinases of family 18 (also known as classes III or V) groups a variety of proteins:
			a) Chitinases from:



		3	363
Pfam	Prosite	Full Name	Description
			- Prokaryotes such as Alteromonas, Bacillus, Serratia, Streptomyces, etc Plants such as Arabidopsis, cucumber, bean, tobacco, etc Fungi such as Aphanocladium, Rhizopus, Saccharomyces, etc Nematode (Brugia malayi) Insects (Manduca sexta) Baculoviruses (Autographa Californica Nuclear Polyhedrosis virus).
			b) Other proteins:
			- Hevamine, a rubber tree protein with chitinase and lysozyme activities. - Kluyveromyces lactis killer toxin alpha subunit, which acts as a chitinase. - Flavobacterium and Streptomyces endo-beta-N-acetylglucosaminidases (EC 3.2. 1.96). - Mammalian di-N-acetylchitobiase which is involved in the degradation of asparagine-linked glycoproteins. - Human cartilage glycoprotein Gp-39. - Jack bean concanavalin B (conB), a protein that has lost its catalytic activity.
			Site directed mutagenesis experiments [3] and crystallographic data [4,5] have shown that a conserved glutamate is involved in the catalytic mechanism and probably acts as a proton donor. This glutamate is at the extremity of the best conserved region in these proteins.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIVMFY]-[DN]-G-[LIVMF]-[DN]-[LIVMF]-[DN]-x-E [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for conB which has a Gln instead of the active site Glu. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Neuhaus JM. jean-marc.neuhaus@bota.unine.ch
			Henrissat B. bernie@afmb.cnrs-mrs.fr
			Last update November 1997 / Text revised. References [1] Flach J., Pilet PE., Jolles P. Experientia 48:701-716(1992).
			[2] Henrissat B. Biochem. J. 280:309-316(1991).
			[3] Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H., Uchida M., Tanaka H. J. Biol. Chem. 268:18567-18572(1993).
			[4] Perrakis A., Tews I., Dauter Z., Oppenheim A.B., Chet I., Wilson K.S., Vorgias C.E. Structure 2:1169-1180(1994).
			[5] van Scheltinga A.C.T., Kalk K.H., Beinterna J.J., Dijkstra B.W. Structure 2:1181-1189(1994).

Pfam	Prosite	Full Name	864 Description
	TOORG	T di (vaine	Description
			[E1] http://www.expasy.ch/cgi-bi/lists?glycosid.txt
01-11			
Choline_kinase		Choline/ethanolamine	Accession number: PF01633
		kinase	Definition: Choline/ethanolamine kinase Author: Bateman A
			Author: Bateman A Alignment method of seed: Clustalw
			Source of seed members: Pfam-B 1165 (release 4.1)
			Gathering cutoffs: 25 25
			Trusted cutoffs: 242.90 242.90
			Noise cutoffs: -85.90 -85.90
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 98175949
			Reference Title: Expression, purification, and characterization of choline
			Reference Title: kinase, product of the CKI gene from
			Saccharomyces
			Reference Title: cerevisiae.
			Reference Author: Kim KH, Voelker DR, Flocco MT, Carmar
			GM;
	[i		Reference Location: J Biol Chem 1998;273:6844-6852.
			Database Reference INTERPRO; IPR002573;
			Comment: Choline kinase catalyses the committed
			step in the synthesis of Comment: phosphatidylcholine by the CDP-choline
			Comment: phosphatidylcholine by the CDP-choline pathway [1].
			Number of members: 22
Chorion		Chorion protein	Accession number: PF01723
			Definition: Chorion protein
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw Source of seed members: Pfam-B_1914 (release 4.1)
			Gathering cutoffs: -46 -46
			Trusted cutoffs: -43.70 -43.70
			Noise cutoffs: -49.00 -49.00
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 95333194 Reference Title: 95333194 Sequence analysis of a small early chorion
			Reference Title: Sequence analysis of a small early chorion gene subfamily
			Reference Title: interspersed within the late gene locus in
			Bombyx mori.
			Reference Author: Kravariti L, Lecanidou R, Rodakis GC;
			Reference Location: J Mol Evol 1995;41:24-33.
			Reference Number: [2]
			Reference Medline: 86313609 Reference Title: Evolution of the silk moth chorion gene
			Reference Title: Evolution of the silk moth chorion gene superfamily: gene
			Reference Title: families CA and CB.
			Reference Author: Lecanidou R, Rodakis GC, Eickbush TH,
			Kafatos FC;
			Reference Location: Proc Natl Acad Sci U S A 1986;83:6514-6518.
			Database Reference INTERPRO; IPR002635;
			Database reference: PFAMB; PB009425;
			Comment: This family consists of the chorion
			superfamily proteins
	ļ		Comment: classes A, B, CA, CB and high-cysteine HCB from silk,
			Comment: gypsy and polyphemus moths.
			Comment: The chorion proteins make up the moths
			egg shell a complex
			Comment: extracellular structure [2].
	ĺ		Number of members: 35
		Chorismate mutase	Accession number: PF01817
Chorismate mut	T I		
norismate_mut	ľ		Definition: Chorismate mutase
Chorismate_mut	ļ		Definition: Chorismate mutase Author: Bateman A



	_	8	65
Pfam	Prosite	Full Name	Description
			Source of seed members: PSI-BLAST 1ecm
			Gathering cutoffs: 5 5
			Trusted cutoffs: 5.10 5.10
			Noise cutoffs: -19.90 -19.90
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 95062155
			Reference Title: The crystal structure of allosteric chorismate mutase at
			Reference Title: 2.2-A resolution.
			Reference Author: Xue Y, Lipscomb WN, Graf R,
			Schnappauf G, Braus G;
			Reference Location: Proc Natl Acad Sci U S A
			1994;91:10814-10818.
			Reference Number: [2]
			Reference Medline: 98307941
			Reference Title: Tyrosine and tryptophan act through the
			same binding site
			Reference Title: at the dimer interface of yeast chorismate
			mutase.
			Reference Author: Schnappauf G, Krappmann S, Braus GH; Reference Location: J Biol Chem 1998;273:17012-17017.
			Reference Location: J Biol Chem 1998;273:17012-17017. Reference Number: [3]
			Reference Number: [3] Reference Medline: 98165805
			Reference Title: So roscos Chorismate mutase-prephenate
			dehydratase from Escherichia
			Reference Title: coli. Study of catalytic and regulatory
			domains using
			Reference Title: genetically engineered proteins.
			Reference Author: Zhang S, Pohnert G, Kongsaeree P,
			Wilson DB, Clardy J,
			Reference Author: Ganem B;
			Reference Location: J Biol Chem 1998;273:6248-6253.
			Database Reference: SCOP; 1csm; fa; [SCOP-USA][CATH-
]		PDBSUM]
			Database Reference INTERPRO; IPR002701;
			Database Reference PDB; 1ecm B; 6; 89;
			Database Reference PDB; 1ecm A; 5; 89;
	ł		Database Reference PDB; 1csm A; 133; 162; PDB; 3csm A; 133; 243;
			Database Reference PDB; 3csm B; 133; 243;
			Database Reference PDB; 4csm A; 133; 243;
. 35-			Database Reference PDB; 4csm B; 133; 243;
			Database Reference PDB; 5csm A; 133; 243;
			Database Reference PDB; 2csm A; 133; 246;
			Comment: Chorismate mutase EC:5.4.99.5 catalyses
			the conversion of
			Comment: chorismate to prephenate in the pathway of
			tyrosine and
			Comment: phenylalanine biosynthesis. This enzyme is
	1		negatively
			Comment: regulated by tyrosine, tryptophan and
			phenylalanine [2,3]. Number of members: 28
			Number of members. 20
CNI hudrolooo	PDOC00712;	Nitrilases / cyanide	Nitrilases (EC 3.5.5.1) are enzymes that convert nitriles into
CN_hydrolase	PDOC00712;	hydratase signatures;	their
ľ	FD0000943	Uncharacterized protein	corresponding acids and ammonia. They are widespread in
		family UPF0012	microbes as well as in
		signature	plants where they convert indole-3-acetonitrile to the hormone
			indole-3-
			acetic acid. A conserved cysteine has been shown [1,2] to be
			essential for
			enzyme activity; it seems to be involved in a nucleophilic
			attack on the
	1		nitrile carbon atom.
	0		Consider hydrotope (EC 4.04.00) convents LICAL to formacide In
	1		Cyanide hydratase (EC 4.2.1.66) converts HCN to formamide. In
			phytopathogenic fungi, it is used to avoid the toxic effect of cyanide released by
			wounded
			plants [3]. The sequence of cyanide hydrolase is evolutionary
			related to that



Prosite Prosite Full Name Description of nitriliases. Yeast hypothetical proteins YIL164c and YIL165c also beld this family. As signature patterns for these enzymes, we selected two conserved regions. The first is located in the N-terminal section while the second, which contains the active site cysteine, is located in the central section of pattern(s) and/or profile(s) Description of pattern(s) and/or profile(s) Consensus pattern G-x(2)-[LIVMFY](2)-x-[IF]-x-E-x(2)-[LIVIY-P-Sequences known to belong to this class detected by the pattl. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST [LIVMFYS]-x-[KR] [C is the active site residue] Sequences known to belong to this class detected by the pattl. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yamada J. Biol. Chem. 267:20746-20751(1992). [3] Wang P., Vanetten H.D. Biochem. Biophys. Res. Commun. 187:1048-1054(1992).	
of nitrilases. Yeast hypothetical proteins YIL164c and YIL165c also belot this family. As signature patterns for these enzymes, we selected two conserved regions. The first is located in the N-terminal section while the second, which contains the active site cysteine, is located in the central set in the contains the active site cysteine, is located in the central set in the contains the active site cysteine, is located in the central set in the contains the active site cysteine, is located in the central set in the contains the active site cysteine, is located in the central set in the contains the active site residue in the central set in the contains a set in the contains a set in the contains the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains a set in the contains and the contains a set in t	
this family. As signature patterns for these enzymes, we selected two conserved regions. The first is located in the N-terminal section while the second, which contains the active site cysteine, is located in the central set of pattern (s) and/or profile(s) Consensus pattern G-x(2)-[LIVMFY](2)-x-[iF]-x-E-x(2)-[LIVIY-P] Sequences known to belong to this class detected by the pall. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST [LIVMFYS]-x-[KR] [C is the active site residue] sequences known to belong to this class detected by the pall. Other sequences in the active site residue sequences known to belong to this class detected by the pall. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751(1992).	
conserved regions. The first is located in the N-terminal section while the second, which contains the active site cysteine, is located in the central set Description of pattern(s) and/or profile(s) Consensus pattern G-x(2)-[LIVMFY](2)-x-[IF]-x-E-x(2)-[LIVI Y-P Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST [LIVMFYS]-x-[KR] [C is the active site residue] Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751(1992).	ng to
Second, which contains the active site cysteine, is located in the central secondaries the active site cysteine, is located in the central secondaries the active site cysteine, is located in the central secondaries the active site cysteine, is located in the central secondaries the active site cysteine, is located in the central secondaries the active site cysteine part of the cysteine sequence and the cysteine sequence and the cysteine sequence and the cysteine sequence and the cysteine sequence and text revised. The cysteine sequence are sequenced as the cysteine sequence and text revised. The cysteine sequence are sequenced as the cysteine sequence and text revised. The cysteine sequence are sequenced as the cysteine sequence are sequenced as the cysteine sequence and text revised. The cysteine sequence are sequenced as the cysteine sequence are sequenced as the cysteine sequence and text revised. The cysteine sequence are sequenced as the cysteine sequence are sequenced.	,
Consensus pattern G-x(2)-[LIVMFY](2)-x-[iF]-x-E-x(2)-[LIVIY-P] Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST [LIVMFYS]-x-[KR] [C is the active site residue] Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yamada J. Biol. Chem. 267:20746-20751(1992).	
Consensus pattern G-x(2)-[LIVMFY](2)-x-[IF]-x-E-x(2)-[LIVI Y-P Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST [LIVMFYS]-x-[KR] [C is the active site residue] Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751(1992).	ction.
Y-P Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST [LIVMFYS]-x-[KR] [C is the active site residue] Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251 (1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751 (1992). [3] Wang P., Vanetten H.D.	
Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST [LIVMFYS]-x-[KR] [C is the active site residue] Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751(1992). [3] Wang P., Vanetten H.D.	1]-x-G-
Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST [LIVMFYS]-x-[KR] [C is the active site residue] Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751(1992). [3] Wang P., Vanetten H.D.	attern
[LIVMFYS]-x-[KR] [C is the active site residue] Sequences known to belong to this class detected by the p ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251 (1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751 (1992). [3] Wang P., Vanetten H.D.	
Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751(1992). [3] Wang P., Vanetten H.D.	:
November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751(1992). [3] Wang P., Vanetten H.D.	
Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251 (1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751 (1992). [3] Wang P., Vanetten H.D.	
Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yama J. Biol. Chem. 267:20746-20751(1992). [3] Wang P., Vanetten H.D.	
Wang P., Vanetten H.D.	ıda H.
The following uncharacterized proteins have been shown [1 share regions of similarities:] to
- Yeast chromosome X hypothetical protein YJL126w Yeast chromosome XII hypothetical protein YLR351c Fission yeast hypothetical protein SpAC26A3.11 Escherichia coli hypothetical protein ybeM Bacillus subtilis hypothetical protein yhcX Mycobacterium tuberculosis hypothetical protein MtCY20G9.06c Synechocystis strain PCC 6803 hypothetical protein sll06	
- A Pseudomonas fluorescens hypothetical protein in pqqF 5'region. - A Staphylococcus hypothetical protein in agr operon.	
Except for yhcX which is larger, these are protein of about 35 Kd which	
contain, in their central section, a well conserved region centered on a cysteine residue.	
Description of pattern(s) and/or profile(s)	
Consensus pattern [GTA]-x(2)-[IVT]-C-Y-D-[LIVM]-x-F-P-x(Sequences known to belong to this class detected by the p ALL.	
Other sequence(s) detected in SWISS-PROT NONE. Last update	
November 1997 / First entry. References	

Ω	6	7
0	Ю	- /

Pfam	Prosite	Full Name	Description
			[1]
			Bairoch A. Unpublished observations (1995).
			Oripublished observations (1995).
CorA		CorA-like Mg2+	Accession number: PF01544
		transporter protein	Definition: CorA-like Mg2+ transporter protein Author: Bateman A
			Author: Bateman A Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_944 (release 4.0)
			Gathering cutoffs: -62 -62
	1		Trusted cutoffs: -5.90 -5.90 Noise cutoffs: -86.20 -86.20
			Noise cutoffs: -86.20 -86.20 HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 98448512 Reference Title: The CorA magnesium transporter gene
			Reference Title: The CorA magnesium transporter gene family.
			Reference Author: Kehres DG, Lawyer CH, Maguire ME;
			Reference Location: Microb Comp Genomics 1998;3:151-169.
			Reference Number: [2] Reference Medline: 99003207
			Reference Title: The CorA Mg2+ transport protein of
			Salmonella typhimurium.
			Reference Title: Mutagenesis of conserved residues in the third membrane
			Reference Title: domain identifies a Mg2+ pore.
			Reference Author: Smith RL, Szegedy MA, Kucharski LM.
			Walker C, Wiet RM,
			Reference Author: Redpath A, Kaczmarek MT, Maguire ME; J Biol Chem 1998;273:28663-28669.
	Ì		Database Reference INTERPRO; IPR002523;
			Database reference: PFAMB; PB041399;
			Comment: The CorA transport system is the primary Mg2+ influx system of Salmonella
			Comment: typhimurium and Escherichia coli. CorA is
			virtually ubiquitous in the
			Comment: Bacteria and Archaea. There are also
			eukaryotic relatives of this protein Number of members; 25
Cys_knot	PDOC00234	Glycoprotein hormones beta chain signatures	Glycoprotein hormones [1,2] (or gonadotropins) are a family of proteins which
		beta criairi signatures	include the mammalian hormones follitropin (FSH), lutropin
			(LSH), thyrotropin
			(TSH) and chorionic gonadotropin (CG), as well as at least two
			forms of fish gonadotropins. All these hormones consist of two glycosylated
			chains (alpha
			and beta). In mammalian gonadotropins, the alpha chain is
			identical in the
			four types of hormones but the beta chains, while homologous, are different.
			The bots chains are pretains of should do an discussion of
			The beta chains are proteins of about 100 to 140 amino acid residues which
			contain twelve conserved cysteines all involved in disulfide
			bonds [3], as
			shown in the following schematic representation.
			+
			+
			+-
			xxxCxxxxxxCxCxCxCxCxCxCxxCxxCxxCxxCxxCx
			xxxCxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
			+-
			+ +
	1		1
			'C': conserved cysteine involved in a disulfide bond.



			368
Pfam:	Prosite	Full Name	Description
			We have developed two patterns for these hormones. The first one, located in the N-terminal section, is a region which has been said to be involved in the association between the two chains of the hormones. The second pattern consists of a cluster of five conserved cysteines in the C-terminal section.
			Description of pattern(s) and/or profile(s)
			Consensus pattern C-[STAGM]-G-[HFYL]-C-x-[ST] [The two C's are involved in disulfide bonds] Sequences known to belong to this class detected by the pattern ALL, except for rat beta-FSH which has Glu in position 2 of the pattern. Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern [PA]-V-A-x(2)-C-x-C-x(2)-C-x(4)-[STD]-[DEY]-C-x(6,8)- [PGSTAVM]-x(2)-C [The five C's are involved in disulfide
			bonds] Sequences known to belong to this class detected by the pattern ALL, except for 5 sequences. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Lapthorn A. adrian@chem.gla.ac.uk
			Last update July 1998 / Patterns and text revised. References [1] Pierce J.G., Parsons T.F. Annu. Rev. Biochem. 50:465-495(1981).
	ı.		[2] Stockell Hartree A., Renwick A.G.C. Biochem. J. 287:665-679(1992).
			[3] Lapthorn A.J., Harris D.C., Littlejohn A., Lustbader J.W., Canfield R.E., Machin K.J., Morgan F.J., Isaacs N.W. Nature 369:455-461(1994).
cytochrome_b_C	PDOC00171	Cytochrome b/b6 signatures	In the mitochondrion of eukaryotes and in aerobic prokaryotes, cytochrome b is a component of respiratory chain complex III (EC 1.10.2.2) - also known as the bc1 complex or ubiquinol-cytochrome c reductase. In plant chloroplasts and cyanobacteria, there is a analogous protein, cytochrome b6, a component of the plastoquinone-plastocyanin reductase (EC 1.10.99.1), also known as the b6f complex. Cytochrome b/b6 [1,2] is an integral membrane protein of approximately 400 amino acid residues that probably has 8 transmembrane segments. In plants and cyanobacteria, cytochrome b6 consists of two subunits encoded by the petB and petD genes. The sequence of petB is colinear with the N-
			terminal part of mitochondrial cytochrome b, while petD corresponds to the C-terminal part. Cytochrome b/b6 non-covalently binds two heme groups, known as b562 and b566. Four conserved histidine residues are postulated to be the ligands of the iron atoms of these two heme groups.



			69
Pfam	Prosite		Description Apart from regions around some of the histidine heme ligands, there are a few conserved regions in the sequence of b/b6. The best conserved of these regions
			includes an invariant P-E-W triplet which lies in the loop that separates the fifth and sixth transmembrane segments. It seems to be important for electron
			transfer at the ubiquinone redox site - called Qz or Qo (where o stands for outside) - located on the outer side of the membrane.
			A schematic representation of the structure of cytochrome b/b6 is shown below.
			+Fe-b566 -+
			<> <cytochrome-b> <cytochrome-b6-petb><cytochrome-b6-petd></cytochrome-b6-petd></cytochrome-b6-petb></cytochrome-b>
			We developed two signature patterns for cytochrome b/b6. The first includes the first conserved histidine of b/b6, which is a heme b562 ligand; the second includes the conserved PEW triplet.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H [H is a heme b562 ligand] Sequences known to belong to this class detected by the pattern ALL, except for 5 sequences. Other sequence(s) detected in SWISS-PROT 15.
			Consensus pattern P-[DE]-W-[FY]-[LFY](2) Sequences known to belong to this class detected by the pattern ALL, except for Odocoileus hemionus (mule deer) and Paramecium tetraurelia cytochrome b. Other sequence(s) detected in SWISS-PROT 1. Last update November 1995 / Patterns and text revised.
			References [1] Howell N. J. Mol. Evol. 29:157-169(1989).
			[2] Esposti M.D., de Vries S., Crimi M., Ghelli A., Patarnello T., Meyer A. Biochim. Biophys. Acta 1143:243-271(1993).
cytochrome_b_N	PDOC00171	Cytochrome b/b6 signatures	In the mitochondrion of eukaryotes and in aerobic prokaryotes, cytochrome b is a component of respiratory chain complex III (EC 1.10.2.2) - also known as the
			bc1 complex or ubiquinol-cytochrome c reductase. In plant chloroplasts and cyanobacteria, there is a analogous protein, cytochrome b6, a component of the plastoquinone-plastocyanin reductase (EC 1.10.99.1), also known as the b6f complex.
			Cytochrome b/b6 [1,2] is an integral membrane protein of approximately 400 amino acid residues that probably has 8 transmembrane segments. In plants and



			870
Plam	Prosite	Full Name	
			<pre>xxxxx</pre>
			Description of pattern(s) and/or profile(s) Consensus pattern [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H [H is a heme b562 ligand] Sequences known to belong to this class detected by the pattern ALL, except for 5 sequences. Other sequence(s) detected in SWISS-PROT 15. Consensus pattern P-[DE]-W-[FY]-[LFY](2)
			Sequences known to belong to this class detected by the pattern ALL, except for Odocoileus hemionus (mule deer) and Paramecium tetraurelia cytochrome b. Other sequence(s) detected in SWISS-PROT 1. Last update November 1995 / Patterns and text revised. References [1] Howell N. J. Mol. Evol. 29:157-169(1989).
cytochrome_c	PDOC00169	Cytochrome c family heme-binding site signature	Esposti M.D., de Vries S., Crimi M., Ghelli A., Patarnello T., Meyer A. Biochim. Biophys. Acta 1143:243-271(1993). In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus



			0 / 1
Pfam	Prosite	Full Name	Description
			sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of
			the two axial ligands of the heme iron. This arrangement is
			shared by all
:			proteins known to belong to cytochrome c family, which presently includes
			cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f
			and reaction
			center cytochrome c.
			Description of pattern(s) and/or profile(s)
			Consensus pattern C-{CPWHF}-{CPWR}-C-H-{CFYW}
			Sequences known to belong to this class detected by the pattern
			ALL, except for four cytochrome c's which lack the first thioether bond.
			Other sequence(s) detected in SWISS-PROT 454.
			Note: some sideshume als have more than a single haved hame
			Note: some cytochrome c's have more than a single bound heme group c4 has 2, c7 has 3, c3 has 4, the reaction center has 4, and
			cc3/Hmc has 16!
			Last update
			June 1992 / Text revised. References
			[1]
			Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).
			1710g. Biophys. Moi. Biol. 43.1-30(1963).
DAHP_synth_2		Class-II DAHP	Members of this family are aldolase enzymes that catalyse the
		synthetase family	first step of the shikimate pathway. These polypeptides can be useful in the synthesis of aromatic
			compounds, such as amino acids, antibiotics, secondary
			metabolites, etc. Such synthesis can occur either in vitro or in
			vivo.
Dala_Dala_ligas		D-ala D-ala ligase	Accession number: PF01820
			Definition: D-ala D-ala ligase
			Author: Bateman A Alignment method of seed: Clustalw
			Source of seed members: PSI-BLAST 2din
			Gathering cutoffs: 25 25
			Trusted cutoffs: 44.90 26.60 Noise cutoffs: 21.50 18.90
	ŀ		HMM build command line: hmmbuild -f HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			Reference Number: [1] Reference Medline: 97207065
			Reference Title: D-alanine:D-alanine ligase: phosphonate
			and phosphinate Reference Title: intermediates with wild type and the Y216F
			mutant.
			Reference Author: Fan C, Park IS, Walsh CT, Knox JR;
			Reference Location: Biochemistry 1997;36:2531-2538. Database Reference: SCOP: 2dln; fa: ISCOP-USAI[CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR000291; Database Reference PDB: 1iov : 3: 303:
			Database Reference PDB; 1iov; 3; 303; Database Reference PDB; 1iov; 3; 303;
			Database Reference PDB; 2dln; 3; 303;
			Comment: This family contains D-alanineD-alanine ligase enzymes EC:6.3.2.4.
			Number of members: 80
D.150	 	Dir. La L	
DHPS	PDOC00630	Dihydropteroate synthase signatures	All organisms require reduced folate cofactors for the synthesis of a variety
		Sy. miceo orginatareo	of metabolites. Most microorganisms must synthesize folate de
			novo because
			they lack the active transport system of higher vertebrate cells which allows
	1	į .	l l
			these organisms to use dietary folates. Enzymes that are
			these organisms to use dietary folates. Enzymes that are involved in the biosynthesis of folates are therefore the target of a variety of



		-	372
Pfam	Prosite	Full Name	Description
			antimicrobial agents such as trimethoprim or sulfonamides.
		•	
			Dihydropteroate synthase (EC 2.5.1.15) (DHPS) catalyzes the condensation of
			6-hydroxymethyl-7,8-dihydropteridine pyrophosphate to para-
			aminobenzoic acid to form 7,8-dihydropteroate. This is the second step in the
			three steps pathway leading from 6-hydroxymethyl-7,8-dihydropterin to 7,8-
			dihydrofolate.
			DHPS is the target of sulfonamides which are substrates analog that compete
			with para-aminobenzoic acid.
			Bacterial DHPS (gene sul or folP) [1] is a protein of about 275 to
			315 amino
			acid residues which is either chromosomally encoded or found on various
			antibiotic resistance plasmids. In the lower eukaryote
			Pneumocystis carinii, DHPS is the C-terminal domain of a multifunctional folate
			synthesis enzyme
			(gene fas) [2].
			We developed two signature patterns for DHPS, the first signature is located
			in the N-terminal section of these enzymes, while the second
			signature is located in the central section.
			Todated in the central section.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIVM]-x-[AG]-[LIVMF](2)-N-x-T-x-D-S-F-x-D-
			x-[SG] Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern [GE]-[SA]-x-[LIVM](2)-D-[LIVM]-G-[GP]-x(2)- [STA]-x-P
			Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update
			November 1997 / Patterns and text revised. References
			[1]
			Slock J., Stahly D.P., Han CY., Six E.W., Crawford I.P. J. Bacteriol. 172:7211-7226(1990).
			 [2]
			Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves
			C.J. Gene 112:213-218(1992).
Di la vina e - !	DD0000700	Dobuduooviassasiassas	
DHquinase_I	PDOC00788	Dehydroquinase class I active site	3-dehydroquinate dehydratase (EC 4.2.1.10), or dehydroquinase, catalyzes the
			conversion of 3-dehydroquinate into 3-dehydroshikimate. It is the third step
			in the shikimate pathway for the biosynthesis of aromatic amino
			acids from chorismate. Two classes of dehydroquinases exist, known as
			types I and II. The
			best studied type I enzyme is from Escherichia coli (gene aroD) and related
			bacteria where it is a homodimeric protein of a chain of about 250
			residues. In fungi, dehydroquinase is part of a multifunctional enzyme
			which catalyzes
			five consecutive steps in the shikimate pathway. In aroD, it has been shown
			been shown

Pfam.	Prosite	Full Name	Description
1 1411	FIOSRE	T UII IVAIIIE	[1] that a histidine is involved in the catalytic mechanism; we
			used the
			region around this residue as a signature pattern.
			- San Allanda Panalana
			Description of pattern(s) and/or profile(s)
			Consensus pattern D-[LIVM]-[DE]-[LIVMN]-x(18,20)-[LIVM](2)-x-[SC]-[NHY]-H- [DN] [H is the active site residue] Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Last update December 1999 / Pattern and text revised.
			References
			Deka R.K., Kleanthous C., Coggins J.R. J. Biol. Chem. 267:22237-22242(1992).
Diphthamide syn		Putative diphthamide	Accession number: PF01866
		synthesis protein	Definition: Putative diphthamide synthesis protein
			Author: Enright A, Ouzounis C, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Enright A Gathering cutoffs: 25 25
			Trusted cutoffs: 44.90 44.90
			Noise cutoffs: -174.70
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 96183112
			Reference Title: A cDNA from the ovarian cancer critical region of deletion
			Reference Title: on chromosome 17p13.3.
			Reference Author: Phillips NJ, Zeigler MR, Deaven LL;
			Reference Location: Cancer Lett 1996;102:85-90.
			Reference Number: [2]
			Reference Medline: 94010339
			Reference Title: Diphthamide synthesis in Saccharomyces cerevisiae:
			Reference Title: structure of the DPH2 gene.
			Reference Author: Mattheakis LC, Sor F, Collier RJ;
			Reference Location: Gene 1993;132:149-154.
			Database Reference INTERPRO; IPR002728;
	•		Comment: Swiss:Q16439 is a candidate tumour
			suppressor gene [1]. DPH2 from Comment: veast Swiss:P32461 [2], which confers
			Comment: yeast Swiss:P32461 [2], which confers resistance to diphtheria toxin
			Comment: has been found to be involved in
			diphthamide synthesis. Diphtheria
			Comment: toxin inhibits eukaryotic protein synthesis by
			ADP-ribosylating Comment: diphthamide, a posttranslationally modified
			Comment: diphthamide, a posttranslationally modified histidine residue present
			Comment: in EF2. The exact function of the members
			of this family is
			Comment: unknown. Number of members: 12
Disintegrin	PDOC00351	Disintegrins signature	Disintegrins [1,2] are snake venom proteins which inhibit fibringen
			interaction with platelet receptors expressed on the glycoprotein IIIb-IIIa
			complex. They act by binding to the integrin glycoprotein IIb-IIIa receptor on
			the platelet surface and inhibit aggregation induced by ADP,
			thrombin, platelet-activating factor and collagen.
			Disintegrins are peptides of about 70 amino acid residues that contain many
			cysteines all involved in disulfide bonds [3]. Disintegrins contain
			an Arg-



	8 / 4
Pfam Prosite Full Name	
	Gly-Asp (RGD) sequence, a recognition site of many adhesion
	proteins. The RGD sequence of disintegrins is postulated to interact with the
	glycoprotein IIb-
	Illa complex.
	The sequences of disintegrins from different snake species are known. These
	proteins are known as: albolabrin, applagin, barbourin,
	batroxostatin,
	bitistatin, echistatin, elegantin, eristicophin, flavoridin, halysin,
	kistrin,
	tergeminin and triflavin.
	Some other proteins are known to contain a disintegrin domain:
	- Some snake venom zinc metalloproteinases [4] consist of an
	N-terminal catalytic domain fused to a disintegrin domain. Such is the
	case for trimerelysin I (HR1B), atrolysin e (Ht-e) and trigramin. It has
	been suggested that these proteinases are able to cleave
	themselves from the
	disintegrin domains and that the latter may arise from such
	a post-
	translational processing. - The beta-subunit of guinea pig sperm surface protein PH30 [5].
	PH30 is a
	protein involved in sperm-egg fusion. The beta subunit
	contains a
	disintegrin at the N-terminal extremity Mammalian epididymial protein 1 (EAP I) [6]. EAP I is
	associated with the
	sperm membrane and may play a role in sperm maturation.
	Structurally, EAP I
	consists of an N-terminal domain, followed by a zinc metalloproteinase
	domain, a disintegrin domain, and a large C-terminal domain that
	contains a
	transmembrane region.
	The schematic representation of the structure of a typical
	disintegrin is
	shown below:
	++
	xxxxxCxCxxxxxxxCCxxxxxCxxxxCxxxxCCxxxCxxxx
	xxxxxxCxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
	'C': conserved cysteine involved in a disulfide bond. '*': position of the pattern.
	As a signature pattern for disintegrins, we selected a
	conserved central
	region that contains five of the cysteines involved in disulfide
	bonds.
	Description of pattern(s) and/or profile(s)
	Consensus pattern C-x(2)-G-x-C-C-x-[NQRS]-C-x-[FM]-x(6)-C-
	[RK] Sequences known to belong to this class detected by the pattern
	ALL. Other sequence(s) detected in SWISS-PROT NONE.
	Last update December 1992 / Pattern and text revised.
	15000mbor 1002 / 1 attorn and tone rondou.

Pfam	Prosite	Full Name	Description
	1		References
			[1] Williams J., Rucinski B., Holt J., Niewiarowski S. Biochim. Biophys. Acta 1039:81-89(1990).
			[2] Dennis M.S., Henzel W.J., Pitti R.M., Lipari M.T., Napier M.A., Deisher T.A., Bunting S., Lazarus R.A. Proc. Natl. Acad. Sci. U.S.A. 87:2471-2475(1990).
			[3] Calvete J.J., Schaefer W., Soszka T., Lu W., Cook J.J., Jameson B.A., Niewiarowski S. Biochemistry 30:5225-5229(1991).
			[4] Hite L.A., Fox J.W., Bjarnason J.B. Biol. Chem. Hoppe-Seyler 373:381-385(1992).
			[5] Blobel C.P., Wolfsberg T.G., Turck C.W., Myles D.G., Primakoff P., White J.M. Nature 356:248-252(1992).
			[6] Perry A.C.F., Jones R., Barker P.J., Hall L. Biochem. J. 286:671-675(1992).
DLH		Dienelactone hydrolase family	Accession number: PF01738 Definition: Dienelactone hydrolase family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_757 (release 4.2) Gathering cutoffs: 15 0 Trusted cutoffs: 15.60 3.10 Noise cutoffs: 14.40 14.40 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 90339491 Reference Medline: 90339491 Reference Title: Refined structure of dienelactone hydrolase at 1.8 A. Reference Author: Pathak D, Ollis D; Reference Location: J Mol Biol 1990;214:497-525. Database Reference: SCOP; 1din; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: Database Reference: PFAMB; PB041131; Database reference: PFAMB; PB041131; Database reference: PFAMB; PB041469; Number of members: 42
DNA_mis_repair	PDOC00057	DNA mismatch repair proteins mutL / hexB / PMS1 signature	Mismatch repair contributes to the overall fidelity of DNA replication [1]. It involves the correction of mismatched base pairs that have been missed by the proofreading element of the DNA polymerase complex. The sequence of some proteins involved in mismatch repair in different organisms have been found to be evolutionary related. These proteins are: - Escherichia coli and Salmonella typhimurium mutL protein [2]. MutL is required for dam-dependent methyl-directed DNA repair. - Streptococcus pneumoniae hexB protein [3]. The Hex system is nick directed. - Yeast proteins PMS1 and MLH1 [4]. - Human protein MLH1 [5] which is involved in a form of familial hereditary nonpolyposis colon cancer (HNPCC).



Pfam	Prosite	Full Name	Description
			As a signature pattern for this class of mismatch repair proteins we selected a perfectly conserved heptapeptide which is located in the N-terminal section of these proteins.
			Description of pattern(s) and/or profile(s) Consensus pattern G-F-R-G-E-A-L Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Pattern and text revised. References [1] Modrich P. Annu. Rev. Biochem. 56:435-466(1987). [2] Mankovich J.A., McIntyre C.A., Walker G.C. J. Bacteriol. 171:5325-5331(1989).
			Prudhomme M., Martin B., Mejean V., Claverys JP. J. Bacteriol. 171:5332-5338(1989). [4] Prolla T.A., Christie D., Liskay R.M. Mol. Cell. Biol. 14:407-415(1994). [5] Bronner C.E., Baker S.M., Morrison P.T., Warren G., Smith L.G., Lescoe M.K., Kane M., Earibino C., Lipford J., Linblom A., Tannergard P., Bollag R.J., Godwin A.R., Ward D.C., Nordenskjold M., Fishel R., Kolodner R.D., Liskay R.M. Nature 368:258-261(1994).
DNA_primase_S		DNA primase small subunit	Accession number: PF01896 Definition: DNA primase small subunit Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 198.40 198.40 Noise cutoffs: -120.80 -120.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 91219475 Reference Title: Mutations in conserved yeast DNA primase domains impair DNA Reference Title: replication in vivo. Reference Author: Santocanale C, Reference Author: Lucchini G, Plevani P; Reference Author: Proc Natl Acad Sci U S A 1991;88:3877-3881. Database Reference Comment: DNA primase Comment: fragments in lagging strand DNA synthesis. DNA primase Comment: is a heterodimer of large and small subunits. Number of members: 14
DnaB		DnaB-like helicase	Members of this family are comprise DNA replication enzymes which unwind the helix. Generally, such polypeptide are ATPases which move at the replication fork, disrupting hydrogen bonds. Such proteins are use for DNA replication in vivo and/or in vitro.
DnaJ_C		DnaJ C terminal region	Accession number: PF01556





Yam	Prosite	Full Name	Description
			Definition: DnaJ C terminal region Author: Bashton M, Bateman A
			Author: Bashton M, Bateman A Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_342 (release 4.0)
			Gathering cutoffs: -24 -24
			Trusted cutoffs: -22.60 -22.60
		<u>'</u>	Noise cutoffs: -25.50 -25.50
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 98308847
	1		Reference Title: The J-domain family and the recruitment of
			chaperone power. Reference Author: Kelley WL;
			Reference Location: Trends Biochem Sci 1998;23:222-227.
			Database Reference INTERPRO; IPR002939;
			Database reference: PFAMB; PB013976;
			Comment: This family consists of the C terminal region
			form the DnaJ
			Comment: protein. Although the function of this region
			is unknown, it
			Comment: is always found associated with DnaJ and
			DnaJ_CXXCXGXG.
			Comment: DnaJ is a chaperone associated with the
	l	1	Hsp70 heat-shock
			Comment: system involved in protein folding and renaturation after stress.
			Number of members: 116
			Number of members.
naJ CXXCXGXG	PDOC00553	dnaJ domains signatures	The prokaryotic heat shock protein dnaJ interacts with the
maj_cxxcxdxd	FD0000000	and profile	chaperone hsp70-
		and prome	like dnaK protein [1]. Structurally, the dnaJ protein consists of
			an N-
			terminal conserved domain (called 'J' domain) of about 70
			amino acids, a
			glycine-rich region ('G' domain') of about 30 residues, a centra
			domain containing four repeats of a CXXCXGXG motif ('CRR' domain)
			and a C-terminal
			region of 120 to 170 residues. Such a structure is shown in the
			following
			schematic representation:
			'
	1		++
			N-terminal Gly-R CXXCXGXG C-terminal
			++
			It has been shown [2] that the 'J' domain as well as the 'CRR'
	1		
			domain are also found in other prokaryotic and eukaryotic proteins which are listed
			below.
			below.
			a) Proteins containing both a 'J' and a 'CRR' domain:
			i e e e e e e e e e e e e e e e e e e e
	1		Vesst service MACEA/D It which cooms to be involved in
			- Yeast protein MAS5/YDJ1 which seems to be involved in
			mitochondrial protein
			mitochondrial protein import.
			mitochondrial protein import Yeast protein MDJ1, involved in mitochondrial biogenesis
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding.
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting.
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting. - Yeast protein XDJ1. - Plants dnaJ homologs (from leek and cucumber).
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting. - Yeast protein XDJ1. - Plants dnaJ homologs (from leek and cucumber).
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting. - Yeast protein XDJ1.
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting. - Yeast protein XDJ1. - Plants dnaJ homologs (from leek and cucumber). - Human HDJ2, a dnaJ homolog of unknown function. - Yeast hypothetical protein YNL077w.
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting. - Yeast protein XDJ1. - Plants dnaJ homologs (from leek and cucumber). - Human HDJ2, a dnaJ homolog of unknown function. - Yeast hypothetical protein YNL077w. b) Proteins containing a 'J' domain without a 'CRR' domain:
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting. - Yeast protein XDJ1. - Plants dnaJ homologs (from leek and cucumber). - Human HDJ2, a dnaJ homolog of unknown function. - Yeast hypothetical protein YNL077w. b) Proteins containing a 'J' domain without a 'CRR' domain: - Rhizobium fredii nolC, a protein involved in cultivar-specific
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting. - Yeast protein XDJ1. - Plants dnaJ homologs (from leek and cucumber). - Human HDJ2, a dnaJ homolog of unknown function. - Yeast hypothetical protein YNL077w. b) Proteins containing a 'J' domain without a 'CRR' domain: - Rhizobium fredii nolC, a protein involved in cultivar-specific nodulation
			mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting. - Yeast protein XDJ1. - Plants dnaJ homologs (from leek and cucumber). - Human HDJ2, a dnaJ homolog of unknown function. - Yeast hypothetical protein YNL077w. b) Proteins containing a 'J' domain without a 'CRR' domain: - Rhizobium fredii nolC, a protein involved in cultivar-specific



fam	Prosite		Description
			into the endoplasmic reticulum and the nucleus Yeast protein SIS1, required for nuclear migration during
			mitosis.
			 Yeast protein CAJ1. Yeast hypothetical protein YFR041c.
			- Yeast hypothetical protein YIR004w.
			 Yeast hypothetical protein YJL162c. Plasmodium falciparum ring-infected erythrocyte surface
			antigen (RESA).
			RESA, whose function is not known, is associated with the membrane skeleton
	i		of newly invaded erythrocytes.
			- Human HDJ1.- Human HSJ1, a neuronal protein.
			- Drosophila cysteine-string protein (csp).
			We developed a signature pattern for the 'J' domain, based on conserved
			positions in the C-terminal half of this domain. We also
			developed a pattern for the first two copies of that motif.
			We also
			developed a profile for the 'J' domain.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [FY]-x(2)-[LIVMA]-x(3)-[FYWHNT]-[DENQSA]-
			x-L-x-[DN]-x(3)- [KR]-x(2)-[FYI]
			Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT 5.
			Consensus pattern C-[DEGSTHKR]-x-C-x-G-x-[GK]-[AGSDM]-
			x(2)-[GSNKR]-x(4,6)-C- x(2,3)-C-x-G-x-G Sequences known to belong to this class detected by the pattern
			ALL, except for yeast XDJ1.
			Other sequence(s) detected in SWISS-PROT 8.
			Sequences known to belong to this class detected by the profile ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Note this documentation entry is linked to both a signature pattern
			and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary
			software tools to do so.
			Expert(s) to contact by email Kelley W. kelley@medecine.unige.ch
			Last update
			July 1998 / Patterns and text revised.
			References
			Cyr D.M., Langer T., Douglas M.G. Trends Biochem. Sci. 19:176-181(1994).
			[2]
			Bork P., Sander C., Valencia A., Bukau B.
			Trends Biochem. Sci. 17:129-129(1992).
			[3] Ueguchi C., Kaneda M., Yamada H., Mizuno T.
			Proc. Natl. Acad. Sci. U.S.A. 91:1054-1058(1994).
dNK		Deoxynucleoside kinase	Accession number: PF01712
			Definition: Deoxynucleoside kinase Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_1744 (release 4.1) Gathering cutoffs: 25 25
			Trusted cutoffs: 47.50 47.50

Noise cutoffs: HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild-re-seed 0 HMM Reference Pite: 19728800 Reference Pite: 19728800 Reference Pite: 19728800 Reference Pite: 19728800 Reference Pite: 19728800 Reference Location: 1982800 Reference Dite: 19728800 Reference Number: 1982800 Referen	Pfam	Prosite	Full Name	Description
HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild bate -seed 0 HMM Reference Number: 87286800 Reference Title: 97226800 Reference Title: 97226800 Reference Title: 97226800 Reference Title: 97226800 Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Packet Reference Author: Refe		100000000000000000000000000000000000000	Tour Hambo	Description Noise cutoffs: -5.40 -5.40
HMM build command line: hmmoalibrateseed 0 HMM Reference Number 11 Reference Number 12				
Reference Number: 97236800 Reference Number: 97236800 Reference Title: 97236800 Reference Title: 197236800 Reference Title: 197236800 Reference Author: Reference Location: Reference Location: Reference Number: 198239511 Reference Mainter: 198239511 Reference Mainter: 198239511 Reference Mainter: 198239511 Reference Mainter: 198239511 Reference Location: Pros Natl Acad Sci U S A 1996:93:7256 Reference Location: Pros Natl Acad Sci U S A 1996:93:7256 Reference Location: Pros Natl Acad Sci U S A 1996:93:7256 Reference Location: Pros Natl Acad Sci U S A 1996:93:7256 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823951 Reference Location: 19823121 Reference Location: 19823121 Reference Location: 19823121 Reference Location: 19823121 Reference Location: 19823121 Reference Location: 19823121-Reference Location: 19823121-Reference Location: 19823121-Reference Location: 19823121-Reference Location: 19823121-Reference Location: 19823121-Reference Location: 19823121-Reference Location: 19823121-Reference Location: 19823121-Reference Location: 19823121-Reference Location: 198232495 Reference Location: 198232458 Refere			1	HMM build command line: hmmcalibrateseed 0 HMM
Reference Title: localization of the gene Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Prox Natl Acad Sci U S A 1996;93:7256 (Author) Author: Reference Author: Refere				Reference Number: [1]
Integral membrane Integral memb				
Reference Title: Reference Nathor: Reference Location: Reference Nathor: Reference N	·			and the state of t
Reference Location: Reference Number: Reference				I = .
Reference Location: Reference Namber: Reference Medline: Reference Medline: Reference Name: Reference Medlin				
Reference Number: [2] Reference Number: [2] Reference Title: Cloining and expression of human decoxyguianosine kinase cDNA. Reference Author: Database Reference Comment: This family consists of various decoxyguianosine consists of various decoxyguianosine consists of various decoxyguides consists of various decoxyguides decoxyguidine Comment: Comment: Comment: Comment: Adenosine EC:2.7.1.74, guanosine EC:2.7.1.74, guanosine EC:2.7.1.75, and the production of Comment: and thymidine kinase EC:2.7.1.21 (which also phosphorylates decoxyguidine Comment: and decoxyguidine comment: and decoxyguidine comment: and decoxyguidine decoxyguidine decoxyguidine comment: and decoxyguidine decoxyguidine comment: and decoxyguidine decoxyguidine decoxyguidine decoxyguidine decoxyguidine comment: and decoxyguidine decoxygu				
Reference Title: Cloning and expression of human decoxyguanosine kinases CDNA. Reference Author: Refer				
deoxyguanosine kinase cDNA. Reference Author: Reference Location: 7262. Database Reference Comment: deoxynucleoside kinases Comment: catalyse the production of Comment: deoxynucleoside. Comment: catalyse the production of Comment: deoxynucleoside. Comment: deoxynucleoside by Comment: catalyse the production of Comment: deoxynucleoside. Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside. Comment: deoxynucleoside comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside by Comment: deoxynucleoside kinases by Comment: deoxynucleoside deoxynucleoside deoxynucleoside deoxynu				
Reference Author: Reference Author: Reference Author: Reference Location: 7262. Database Reference Comment: Geoxynucleoside kinases Comment: EC.2.7.1.74, guanosine EC.2.7.1.74 Comment: also phosphorylates deoxynucleoside. Comment: Geoxynucleoside. Comment: Geoxynucleoside. Comment: Geoxynucleoside. Comment: Using ATP and yielding ADP in the process of the production of deoxynucleoside. Comment: Number of members: Delta serrate ligand Author: Porting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 2 525 Trusted cutoffs: 3.40 3.40 Noise				
Reference Location: 7262. Database Reference Comment: This family consists of various deoxynucleoside kinases Comment: Anything Consists of various deoxynucleoside kinases Comment: Anything Consists of various deoxynucleoside kinases Comment: Anything Comment: Anything Consists of various and thymidine kinase EC:2.7.1.74, guanosine EC:2.7.1.173, demonstrate EC:2.7.1.173, demonstrate Comment: Anything Consists of Comment: Anything Comment: C				
Database Reference Comment: deoxynucleoside kinases Comment: Comment: deoxynucleoside comment: Comment: deoxynucleoside comment: Comment: deoxynucleoside. Comment: deoxynucleoside. Comment: deoxynucleoside. Comment: Desta serrate ligand Accession number: Delta serrate ligand Accession number: Delta serrate ligand Author: Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 2.52 Trusted cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: MART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 3.40.3.40 Alignment members and intrinsic signalling activity of their domains in vivo. Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald I; Fitzgerald K, Greenwald				
Comment: This family consists of various deoxynucleoside kinases: Comment: cytidine EC:2.7.1.74, guanosine EC:2.7.1.75 Comment: and thymidine kinase EC:2.7.1.21 (which also phosphorylates deoxynucliene) These enzymes catalyse the production of Comment: and deoxycytosine.) These enzymes catalyse the production of deoxynucleoside. Comment: Using ATP and yielding ADP in the process Number of members: 20 Delta serrate ligand Accession number: PF01414 Definition: Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Aligament method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 25 Trusted cutoffs: 40 04 30 04 30 40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbailbateseed 0 HMM Efference Number: Reference Title: DSL proteins Reference Title: DSL proteins Reference Title: extracellular Reference Title: extracellular Reference Title: network of the protein seed				7262.
Comment: cytidine EC:27.1.74, guanosine EC:27.1.75 EC:27.1.113, adenosine EC:27.1.76 Comment: and thymidine kinase EC:2.7.1.21 (which also phosphorylates decoyurdine Comment: and deoxycytosine.) These enzymes catalyse the production of Comment: deoxynucleoside. Comment: Using ATP and yielding ADP in the process (Admusi Superior of Comment: Using ATP and yielding ADP in the process (Admusi Superior of Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gatthering cutoffs: 25 25 Trusted cutoffs: 43.00 43.00 Noise cutoffs: 3.40 3.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild end with the process of the pr				Database Reference INTERPRO; IPR002624;
Comment: cytldine EC:2.7.1.74, guanosine EC:2.7.1.13, adenosine EC:2.7.1.19 EC:2.7.1.113, adenosine EC:2.7.1.29 Comment: and thymidine kinase EC:2.7.1.21 (which also phosphorylates deoxyurdine Comment: and deoxycytosine.) These enzymes catalyse the production of Comment: deoxynucleoside. Comment: Using ATP and yielding ADP in the process Number of members: 20 Delta serrate ligand Accession number: PF01414 Definition: Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Aligament method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 25 Trusted cutoffs: 40 04 30 40 HMM build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build command line: hmmbuild HMM SEED HMM Build comm				Comment: This family consists of various
EC:2.7.1.113, adenosine EC:2.7.1.76 Comment: and thymidine kinase EC:2.7.1.21 (which also phosphorylates deoxyuridine Comment: and deoxycytosine.) These enzymes catalyse the production of deoxynucleoside. Comment: Using ATP and yielding ADP in the process Number of members: 20 Delta serrate ligand Accession number: PF01414 Definition: Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 34,03.43.00 Noise cutoffs: 34,03.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM proteins Reference Nedline: Reference Rediline: Reference Title: Reference Title: Reference Title: Reference Title: Reference Cutotion: Reference Number: Reference Number: Reference Number: Reference Title: Reference Title: Reference Title: Reference Title: Reference Cutotion				Fig. 1
Comment: and thymidine kinase EC:2.7.1.21 (which also phosphorylates deoxyurdine) Comment: and deoxycytosine.) These enzymes catalyse the production of Comment: deoxynucleoside. Comment: Using ATP and yielding ADP in the process Number: 20 Delta serrate ligand Accession number: PF01414 Definition: Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25.25 Trusted cutoffs: 3.40.3.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HMM seed hmmbuild HM				
also phosphorylates deoxyuridine Comment: catalyse the production of Comment: deoxynucleoside 5'-monophosphate from a deoxynucleotide from a deoxynucleotide 5'-monophosphate from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide from a deoxynucleotide fromanic from a deoxing from a deoxing from and index from a dinum from and index from a dinum from and index from a dinum from and i				
Comment: catalyse the production of Comment: deoxynucleoside. Comment: Momber of members: 20 Delta serrate ligand Accession number: PF01414 Definition: Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 25 Trusted cutoffs: 3,40 3,40 3,00 43,00 Noise cutoffs: 3,40 3,40 Noise cutoffs: 3,40 3,40 Noise cutoffs: 1951 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 1951 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 1951 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 1951 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 1951 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 1951 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 3,40 3,40 Noise cutoffs: 1951 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 1951 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 1951 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 1951 Trusted cutoffs: 1951 Truste				
catalyse the production of Comment: deoxynucleoside 5'-monophosphate from a deoxynucleoside. Comment: Using ATP and yielding ADP in the process Number of members: 20 Delta serrate ligand Accession number: PF01414 Definition: Delta serrate ligand Author: Ponting CP, Schuttz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 25 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 33,00 43,00 Noise cutoffs: 43,00 43,00 Noise cutoffs: 43,00 43,00 Noise cutoffs: 43,00 43,00 Noise cutoffs: 43,00 43,00 Noise cutoffs: 45,00 43,00 Integrald HMM build cemmand line: himmealibrate with the cutoffs: 45,00 43,00 Noise cutoffs: 45,00 43,00 Noise cutoffs: 45,00 44,00 Noise cutoffs: 45,00				Comment: and deoxycytosine.) These enzymes
Delta serrate ligand Accession number:				catalyse the production of
Delta serrate ligand Accession number: PF01414 Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 25 Trusted cutoffs: 43,00 43.00 Noise cutoffs: 34,00 44.00 Noise cutoffs: 34,00 44.00 Noise cutoffs: 34,00 44.00 Noise cutoffs: 34,00 44.00 Noise cutoffs: 49,00 43.00 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: Reference Mille: DSL proteins Reference Title: DSL proteins Reference Title: Nerracellular Reference Title: Nerracellular Reference Author: Reference Title: Nerracellular Reference Number: Reference Title: Nerracellular Reference Title: Nerracellular Reference Title: Nerracellular Reference Title: Nerracellular Reference Title: Nerracellular Reference Title: Specific EGF repeats of Notch mediate Reference Title: Nerracellular Reference Title: Nerracellular Reference Title: Nerracellular Reference Title: Nerracellular Reference Title: Nerracellular Reference Mumber: Reference Title: Specific EGF repeats of Notch mediate Reference Title: Nerracellular Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Nervanis-Tsakonas S, Cell 1991;67:687-699. Secure 1995;268:225-232. SMART; DSL; Reference Author of members: Number of members: Number of members: Number of members: Number of members: PF01988				are in the state of the state o
Delta serrate ligand Accession number: PF01414 Definition: Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 43,00 43,00 Noise cutoffs: 40,00 41,00				•
Delta serrate ligand Accession number: PF01414 Definition: Pointing CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 25 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 3,40 3,40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild pill hMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 96125168 Reference Title: and intrinsic signalling activity of their extracellular Reference Author: Reference Author: Reference Author: Reference Location: Reference Title: interactions with Reference Title: interactions with Reference Title: nervice Author: L, Cherbas P, L, Cherbas P, L, Cherbas P, L, Cherbas P, Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Medline: Reference Medline: Reference Author: Reference Medline: Reference Medline: Reference Author: Reference Medline: Reference Author: Reference Medline: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Medline: Reference Author: Reference SMART: DSL: INTERPRO; IPR001774; 300 UUF125				
Definition: Delta serrate ligand Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 5 Trusted cutoffs: 43.00 43.00 Noise cutoffs: 3, 40, 3, 40 HMM build command line: hmmcalibrate -seed 0 HMM Reference Number: 18 66125168 Reference Medline: Reference Medline: Post proteins Reference Title: Interchangeability of Caenorhabditis elegan DSL proteins Reference Title: And intrinsic signalling activity of their extracellular Reference Title: Reference Number: Reference Number: Reference Number: Reference Number: Reference Medline: Reference Medline: Reference Title: Interactions with Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Number: Referen				Number of members.
Definition: Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMAFT Gathering cutoffs: 25 25 Trusted cutoffs: 43.00 43.00 Noise cutoffs: 3.40 3.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM seference Number: [1] 96125168 Interchangeability of Caenorhabditis elegan DSL proteins Reference Title: and intrinsic signalling activity of their extracellular Reference Author: Reference Author: Reference Number: Reference Number: Reference Number: Reference Title: Interactions with Reference Title: Delta and Serrate: implications for Notch as a Reference Title: Delta and Serrate: implications for Notch as a Reference Author: L, Cherbas P, Reference Author: Reference Author: Reference Author: Reference Number: Reference Medline: Reference Medline: Reference Mumber: Reference Author: Database reference: Database reference: Database Reference: Database Reference Number: Database Reference: Number: Database Reference: Number: Database Reference: Number: Database Reference: Number: PF01988	DSL		Delta serrate ligand	Accession number: PF01414
Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 25 Trusted cutoffs: 43,00 43,00 Noise cutoffs: 3,40 3.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM seed on should line: hmmbuild HMM seed on should line: hmmbuild HMM seed on should line: hmmbuild line: hmmbuild line: hmmbuild line: hmmbuild line: hmmbuild line: hmmbuild line: hmmbuild line: hmmbuild line: hmmbuild line: hmmbuild line: hmmbuild line: hmmbuild line: hm		İ		
Source of seed members: SMART Gathering cutoffs: 43.00 43.00 Noise cutoffs: 3.40 3.40 HMM build command line: hmmcalibrateseed 0 HMM Reference Number: Reference Medline: hiteractions with Reference Title: extracellular Reference Title: extracellular Reference Title: extracellular Reference Author: Reference Location: Reference Medline: Reference Title: extracellular Reference Title: extracellular Reference Title: neteractions with Reference Title: interactions with Reference Title: interactions with Reference Title: Reference Author: L, Cherbas P, Reference Author: L, Cherbas P, Reference Author: Reference Author: Reference Author: L, Cherbas P, Reference Author: L, Cherbas P, Reference Author: L, Cherbas P, Reference Author: Reference Author: Reference Author: L, Cherbas P, Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: L, Cherbas P, Reference Author: Reference Ref				
Gathering cutoffs: 25 25 Trusted cutoffs: 3,00 43,00 Noise cutoffs: 3,40 3,40 HMM build command line: hmmbaild HMM SEED HMMM build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbaild HMM SEED HMM Build command line: hmmbail lile; hiterations with and intrinsic signalling activity of their domains in vivo. Fitzgerald K, Greenwald I; Devlopment 1995;121:4275-4282. [2] Specific EGF repeats of Notch mediate Heference Title: ndmains in vivo. Fitzgerald K, Greenwald I; Devlopment				Alignment method of seed: Manual
Trusted cutoffs: 43.00 43.00 Noise cutoffs: 3.40 3.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild research in the state of the sta				
Noise cutoffs: 3,40 3,40 140 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuilt line: hmmballity of Caenorhabditis elegan doit line: hmmbuilt line: hmmbuilt line: hmmbuilt line: hmmballity of Caenorhabditis elegan doit line: hmmbuilt				
HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED Reference Number: [1] Reference Medline: Reference Title: bSL proteins Reference Title: otherwise and intrinsic signalling activity of their extracellular Reference Title: otherwise and intrinsic signalling activity of their extracellular Reference Title: otherwise and intrinsic signalling activity of their extracellular Reference Title: otherwise and intrinsic signalling activity of their extracellular Reference Title: otherwise and intrinsic signalling activity of their extracellular Reference Author: Reference Author: Reference Location: Reference Medline: Reference Title: Reference Author: Reference Author: Reference Author: Reference Number: Reference Author: ME; Reference Author: ME; Reference Author: ME; Reference Author: ME; Reference Location: Database reference: Database Reference Number of members: Database Reference Number: PF01988 HMM build command line: hmmbufilie: hmmbufilie: hmmbufilie: hmmbufilie: hmmbufilie: hmmbufilie: hmmbufilie: hmmbufilie: hmmbufilie: hmmbufilie: hmmbufilie: helpan and intrinsic signalling activity of their extracellular and intrinsic signalling activity of their extracellular and intrinsic signalling activity of their extracellular and intrinsic signalling activity of their extracellular and intrinsic signalling activity of their extracellular and intrinsic signalling activity of their extracellular and intrinsic signalling and intrinsic signalling activity of their extracellular and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling and intrinsic signalling a				
HMM build command line: hmmcalibrateseed 0 HMM Reference Number: Reference Title: DSL proteins Reference Title: extracellular Reference Title: Reference Author: Reference Number: Reference Number: Reference Number: Reference Number: Reference Title: Reference Medline: Reference Number: Reference Number: Reference Medline: Reference Number: Reference Number: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Author: L, Cherbas P, Reference Author: Reference Number: Reference Number: Reference Author: Reference Number: Reference Author:				
Reference Medline: Reference Title: DSL proteins Reference Title: extracellular Reference Title: neference Title: Reference Author: Reference Number: Reference Number: Reference Title: interactions with Reference Title: Reference Title: neference Title: neference Title: Reference Title: Reference Title: Reference Title: Reference Author: L, Cherbas P, Reference Author: Reference Title: Reference Author: Reference Author: Reference Title: Reference Author: Reference Title: Reference Title: Reference Author: Reference Title: R				HMM build command line: hmmcalibrateseed 0 HMM
Reference Title: DSL proteins Reference Title: extracellular Reference Title: extracellular Reference Title: extracellular Reference Title: extracellular Reference Author: Reference Author: Reference Medline: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Title: Interactions with Reference Title: Interactions with Reference Title: Interactions in vivo. Fitzgerald K, Greenwald I; Development 1995;121:4275-4282. [2] Specific EGF repeats of Notch mediate Delta and Serrate: implications for Notch as a Reference Author: Reference Author: Reference Author: Reference Number: Reference Number: Reference Medline: Reference Medline: Reference Title: Interchangeability of Caenorhabditis elegan and intrinsic signalling activity of their domains in vivo. Fitzgerald K, Greenwald I; Development 1995;121:4275-4282. [2] Specific EGF repeats of Notch mediate Delta and Serrate: implications for Notch as a multifunctional receptor. Refeat and intrinsic signalling activity of their domains in vivo. Fitzgerald K, Greenwald I; Development 1995;121:4275-4282. [2] Specific EGF repeats of Notch mediate Multifunctional receptor. Refear (Cell 1991;67:687-699. [3] Specific EGF repeats of Notch mediate Notch as a Artavanis-Tsakonas S; Cell 1991;67:687-699. [3] Specific EGF repeats of Notch mediate Multifunctional receptor. Reference Medline: Reference Number: Reference Author: Reference Author: Reference Author: Reference Title: Interchangeability of Caenorhabditis elegan domains in vivo. Fitzgerald K, Greenwald I; Development 1995;121:4275-4282. [2] Specific EGF repeats of Notch mediate Notch as a Artavanis-Tsakonas S; Cell 1991;67:687-699. [3] Specific EGF repeats of Notch mediate Notch as a Artavanis-Tsakonas S; Cell 1991;67:687-699. [3] Specific EGF repeats of Notch mediate Notch as a Artavanis-Tsakonas S; Cell 1991;67:687-699. [3] Specific EGF repeats of Notch mediate Notch as a Artavanis-Tsakonas S; Cell 1991;67:687-699. [3] Specific EGF repeats of Notch me				
DSL proteins Reference Title: extracellular Reference Author: Reference Number: Reference Medline: Reference Title: Reference Number: Reference Medline: Reference Title: Reference Title: Reference Medline: Reference Title: Interactions with Reference Title: Reference Title: Interactions with Reference Title: Reference Author: L, Cherbas P, Reference Author: Reference Author: Reference Author: Reference Number: Reference Number: Reference Number: Reference Author: Reference Author: Reference Author: Reference Number: Reference Author: Reference Title: Reference Title: Reference Title: Reference Medline: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Author: Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author:				
Reference Title: extracellular Reference Title: Reference Title: domains in vivo. Reference Author: Reference Location: Reference Mumber: Itle: interactions with Reference Author: Reference Author: L., Cherbas P, Reference Author: Reference Number: Reference Number: Reference Number: Reference Author: L., Cherbas P, Reference Number: Reference Author: NE; Reference Author: ME; Reference Author: ME; Reference Coation: Database Reference Number: Database Reference Number: Number of members: Number of members: Number: PF01988				grammy or a marrier madarile drogario
extracellular Reference Title: Reference Author: Reference Number: Reference Medline: Reference Title: Reference Title: Reference Title: Reference Title: Interactions with Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Author: L, Cherbas P, Reference Author: Reference Number: Reference Number: Reference Number: Reference Author: Reference Number: Reference Author: Reference Title: Reference Author: Reference Author: Reference Title: Reference Author: Reference Author: Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Title: Reference Title: Reference Author: Reference Number: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Number: Reference Author: Reference Author: Reference Author: R				le
Reference Author: Reference Location: Reference Number: Reference Neddline: Reference Medline: Reference Title: Interactions with Reference Author: Reference Author: Reference Author: L, Cherbas P, Reference Author: Reference Number: Reference Number: Reference Neddline: Reference Author: Reference Author: Reference Number: Reference Number: Reference Author: Reference Number: Reference Author: Reference Number: Reference Number: Reference Author: Reference Number: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Number: Reference				
Reference Location: Reference Number: Reference Number: Reference Title: Interactions with Reference Title: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Number: Reference Number: Reference Number: Reference Number: Reference Author: Reference Number: Reference Number: Reference Author: Reference Number: Reference Number: Reference Author: Reference Number: Reference Number: Reference Author: Reference Number: Reference Number: Reference Number: Reference Number: Reference Author: Reference Number: Reference Nu				
Reference Number: Reference Medline: Reference Title: interactions with Reference Title: a Reference Title: neference Title: neference Title: neference Title: neference Title: Reference Author: L, Cherbas P, Reference Author: Reference Location: Reference Number: Reference Medline: Reference Medline: Reference Author: Reference Title: Reference Medline: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Location: Reference Author: ME; Reference Location: Database reference: Database Reference Number of members: Integral membrane Reference Notch signaling. Artavanis-Tsakonas S, Matsuno K, Fortini Science 1995;268:225-232. Science 1995;268:225-232. INTERPRO; IPR001774; 30				
Reference Medline: Reference Title: Interactions with Reference Title: Interactions with Reference Title: Interactions with Reference Title: Interactions with Reference Title: Interactions with Reference Title: Reference Title: Reference Author: L. Cherbas P, Reference Author: Reference Location: Reference Number: Reference Number: Reference Mittle: Reference Mittle: Reference Mittle: Reference Mittle: Reference Mittle: Reference Author: Reference Mittle: Reference Author: Reference Author: Reference Mumber: Reference Author: Reference Mumber: Reference Author: Reference Mumber: Reference Author: Reference Mumber: Reference Author: Reference Mittle: Notch signaling. Artavanis-Tsakonas S, Matsuno K, Fortini Reference Location: Database reference: Database Reference Number of members: Integral membrane Accession number: PF01988				
Reference Title: interactions with Reference Title: a Reference Title: A Reference Title: Reference Author: L, Cherbas P, Reference Author: Reference Location: Reference Medline: Reference Author: Reference Author: Reference Title: Reference Medline: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Medline: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: Reference Medline: Reference Author: ME; Reference Author: ME; Reference Location: Database reference: Database Reference Number of members: MITERPRO; IPR001774; MITERP				
interactions with Reference Title: a Reference Title: Reference Author: L, Cherbas P, Reference Author: Reference Number: Reference Medline: Reference Author: Reference Title: Reference Author: Reference Author: Reference Number: Reference Author: Reference Author: Reference Number: Reference Author: Reference Author: Reference Author: Reference Author: Reference Author: ME; Reference Location: ME; Reference Location: Database reference: Database Reference Number of members: Integral membrane Delta and Serrate: implications for Notch as a multifunctional receptor. Rebay I, Fleming RJ, Fehon RG, Cherbas Cell 1991;67:687-699. [3] Sp232495 Notch signaling. Artavanis-Tsakonas S, Matsuno K, Fortini ME; Reference Location: Science 1995;268:225-232. SMART; DSL; INTERPRO; IPR001774; 30				
a Reference Title: Reference Author: L, Cherbas P, Reference Author: Artavanis-Tsakonas S; Reference Number: Reference Medline: Notch signaling. Reference Author: ME; Reference Location: Database reference: Database Reference Number: Database Reference Number: Database Reference Number: Database Reference Number: Database Reference Number: Database Reference Number: Notch signaling. Reference Location: Science 1995;268:225-232. SMART; DSL; INTERPRO; IPR001774; 30 Integral membrane Accession number: PF01988				interactions with
Reference Title: Reference Author: L, Cherbas P, Reference Author: Reference Author: L, Cherbas P, Reference Location: Reference Number: Reference Medline: Reference Author: Reference Author: Reference Medline: Reference Author: Reference Author: Reference Title: Notch signaling. Artavanis-Tsakonas S; Notch signaling. Artavanis-Tsakonas S, Matsuno K, Fortini ME; Reference Author: ME; Reference Location: Database reference: Database Reference: Number of members: NUF125 Integral membrane Meteronce Title: Reference Author: Artavanis-Tsakonas S, Matsuno K, Fortini Science 1995;268:225-232. SMART; DSL; INTERPRO; IPR001774; 30				Delta and Contate: implications for Noteri as
Reference Author: L, Cherbas P, Reference Author: Reference Author: Artavanis-Tsakonas S; Cell 1991;67:687-699. Reference Number: Reference Medline: Reference Author: Notch signaling. Artavanis-Tsakonas S; Notch signaling. Artavanis-Tsakonas S, Matsuno K, Fortini ME; Reference Author: ME; Reference Location: Database reference: Database Reference: Number of members: Integral membrane Rebay I, Fleming RJ, Fehon RG, Cherbas Artavanis-Tsakonas S, Notch signaling. Artavanis-Tsakonas S, Matsuno K, Fortini Science 1995;268:225-232. SMART; DSL; INTERPRO; IPR001774; 30				
L, Cherbas P, Reference Author: Artavanis-Tsakonas S; Reference Location: Cell 1991;67:687-699. Reference Number: [3] Reference Medline: 95232495 Reference Title: Notch signaling. Reference Author: Artavanis-Tsakonas S, Matsuno K, Fortini ME; Reference Location: Science 1995;268:225-232. Database reference: SMART; DSL; Database Reference Number of members: 30 Integral membrane Accession number: PF01988				
Reference Author: Reference Location: Reference Number: Reference Medline: Reference Title: Reference Author: ME; Reference Location: ME; Reference Location: Database reference: Database Reference: Number of members: Integral membrane Artavanis-Tsakonas S; Cell 1991;67:687-699. [3] Artavanis-Tsakonas S, Matsuno K, Fortini Science 1995;268:225-232. SMART; DSL; INTERPRO; IPR001774; 30				L. Cherbas P
Reference Location: Cell 1991;67:687-699. Reference Number: [3] Reference Medline: 95232495 Notch signaling. Reference Author: ME; Reference Location: Science 1995;268:225-232. Database reference: Database Reference Number of members: Note signaling. Artavanis-Tsakonas S, Matsuno K, Fortini Science 1995;268:225-232. SMART; DSL; INTERPRO; IPR001774; 30 Integral membrane Accession number: PF01988				· · · · · · · · · · · · · · · · · ·
Reference Number: [3] Reference Medline: 95232495 Reference Title: Notch signaling. Reference Author: ME; Reference Location: Science 1995;268:225-232. Database reference: Database Reference Number of members: 30 Integral membrane Reference Number: PF01988				
Reference Title: Notch signaling. Reference Author: Artavanis-Tsakonas S, Matsuno K, Fortini ME; Reference Location: Science 1995;268:225-232. Database reference: SMART; DSL; Database Reference Number of members: 30 DUF125 Integral membrane Reference Title: Notch signaling. Artavanis-Tsakonas S, Matsuno K, Fortini ME; Notch signaling. Artavanis-Tsakonas S, Matsuno K, Fortini ME; NTERPRO; 1980:1774; NOTCH SIGNAL SIGN				Reference Number: [3]
Reference Author: Artavanis-Tsakonas S, Matsuno K, Fortini ME; Reference Location: Science 1995;268:225-232. Database reference: SMART; DSL; Database Reference INTERPRO; IPR001774; Number of members: 30 DUF125 Integral membrane Accession number: PF01988				
ME; Reference Location: Science 1995;268:225-232. Database reference: SMART; DSL; Database Reference INTERPRO; IPR001774; Number of members: 30 DUF125 Integral membrane Accession number: PF01988			!	
Reference Location: Science 1995;268:225-232. Database reference: SMART; DSL; Database Reference INTERPRO; IPR001774; Number of members: 30 Integral membrane Accession number: PF01988				The state of the s
Database reference: SMART; DSL; Database Reference INTERPRO; IPR001774; Number of members: 30 UF125 Integral membrane Accession number: PF01988				le é la la la la la la la la la la la la la
Database Reference INTERPRO; IPR001774; Number of members: 30 UF125 Integral membrane Accession number: PF01988				
Number of members: 30 UF125 Integral membrane Accession number: PF01988				Database Reference INTERPRO; IPR001774;
Accession hamber. Pro1966				
Accession hamber. Pro1966				
Accession hamber. Pro1966	UF125		Integral membrano	Accession number: PEGGGG
Integral membrane protein DUF 125				
				integral membrane protein DOF 125

		8	380
Pfam	Prosite [Full Name	Description
			Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -60 -60 Trusted cutoffs: -57.90 -57.90 Noise cutoffs: -64.60 -64.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95028150 Sequence, mapping and disruption of CCC1, a gene that Reference Title: cross-complements the Ca(2+)-sensitive phenotype of csg1 Reference Author: Fu D, Beeler T, Dunn T; Reference Location: Yeast 1994;10:515-521. Database Reference Comment: This family of predicted integral membrane proteins has no known Comment: function. However it does include Swiss:P47818, that may have a Comment: role in regulating calcium levels [1]. Number of members: 7
DUF25	I I	Domain of unknown function DUF25	Accession number: PF01641 Definition: Domain of unknown function DUF25 Author: Bateman A, Enwright A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1539 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 151.80 151.80 Noise cutoffs: 10.60 10.60 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 20076492 Reference Title: using a conserved RNA structural motif. Reference Author: A; Reference Location: Database Reference Comment: Jabio Chem 1999;274:38147-38154. INTERPRO; IPR002579; This domain has no known function. It is found associated Comment: with the peptide methionine sulfoxide reductase enzymatic Comment: domain PMSR. The domain has two conserved cysteine Comment: and histidines that could suggest and zinc binding site. Comment: The final cysteine is found to be replaced by the rare amino Comment: acid selenocysteine in some members of the family [1]. Number of members: 26
DUF26	1	Domain of unknown function DUF26	Accession number: PF01657 Definition: Domain of unknown function DUF26 Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_980 (release 4.1) Gathering cutoffs: -8 -8 Trusted cutoffs: 6.50 1.40 Noise cutoffs: -17.50 -17.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference INTERPRO; IPR002902; Database reference: PFAMB; PB005223; Comment: This domain has no known function. It is found in serine/threonine Comment: kinases, associated with the Eukaryotic protein kinase domain



Pfam	Prosite	Full Name	Description
			Comment: pkinase. In the 33kDa secretary protein
			Swiss:082551
			Comment: this domain is duplicated. The domain contains four conserved
			Comment: cysteines.
			Number of members: 25
DUF89			Accession number: PF01937
			Definition: Protein of unknown function DUF89
			Author: Enright A, Ouzounis C, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Enright A Gathering cutoffs: 25 25
			Trusted cutoffs: 636.30 636.30
			Noise cutoffs: -142.40 -142.40
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Database Reference INTERPRO; IPR002791;
			Comment: This prokaryotic family has no known
			function. The protein
			Comment: has two closely spaced conserved cysteines
			at its N Comment: terminus and a single conserved cysteine at
			its C terminus.
			Number of members: 5
DUF90			Accession number: PF01938
			Definition: Domain of unknown function DUF90
			Author: Enright A, Ouzounis C, Bateman A
			Alignment method of seed: Clustalw Source of seed members: Enright A
			Gathering cutoffs: 25 0
			Trusted cutoffs: 78.90 10.20
			Noise cutoffs: -0.60 -0.60
			HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Database Reference INTERPRO; IPR002792;
			Comment: This small domain has no known function.
			However it Comment: may perform a nucleic acid binding role
			Comment: may perform a nucleic acid binding role (Bateman A.
			Comment: unpublished observation).
			Number of members: 17
Oynein_light	PDOC00953	Dynein light chain type 1	Dynein is a multisubunit microtubule-dependent motor enzyme
		signature	that acts as the
	1		force generating protein of eukaryotic cilia and flagella. The cytoplasmic
			isoform of dynein acts as a motor for the intracellular retrograde
			motility of
			vesicles and organelles along microtubules. Dynein is composed of a number of
			ATP-binding large subunits, intermediate size subunits and small
			subunits.
			Among the small subunits, there is a family [1,2] of highly
			conserved proteins
			which consist of:
			- Chlamydomonas reinhardtii flagellar outer arm dynein 8 Kd and
			11 Kd light
			chains.
			- Higher eukaryotes cytoplasmic dynein light chain 1.
			- Yeast cytoplasmic dynein light chain 1 (gene DYN2 or SLC1) Caenorhabditis elegans hypothetical dynein light chains M18.2
			and T26A5.9.
			These proteins are have from 80 to 100 amiles saids. A
			These proteins are have from 89 to 120 amino acids. As a signature pattern,
			we selected a highly conserved region.
			Description of nottoro(a) and (as a self-to)
			Description of pattern(s) and/or profile(s)

		8	82
Pfam	Prosite	Full Name	Description
			Consensus pattern H-x-I-x-G-[KR]-x-F-[GA]-S-x-V-[ST]-[HY]-E Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / First entry. References [1] King S.M., Patel-King R.S. J. Biol. Chem. 270:11445-11452(1995). [2] Dick T., Ray K., Salz H.K., Chia W. Mol. Cell. Biol. 16:1966-1977(1996).
elF5_elF2B		Domain found in IF2B/IF5	Accession number: PF01873 Definition: Domain found in IF2B/IF5 Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 233.00 233.00 Noise cutoffs: -56.10 -56.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 96060092 Multidomain organization of eukaryotic guanine nucleotide Reference Title: guanine nucleotide Reference Title: subunits Reference Title: sequence motifs. Reference Author: Reference Location: Database Reference Comment: This region Comment: Swiss:P55010, and Comment: corresponds to the whole of the archaebacterial eIF-2 beta Comment: zinc binding C4 finger. Number of members: 20
elF6		eIF-6 family	This family comprises members exhibiting sequence identity to the eukaryotic translation initiation factor 6. Some members of this family are implicated in protein biosynthesis as a translation initiation factor by binding to the 60s ribosomal subunit and preventing its association with the 40s ribosomal subunit to form the 80s initiation complex. Such activity can play a role in maximal polysome formation and plays an important role in determining free 60s ribosomal subunit content. Polypeptides in this family can optimize amino acid and nitrogen content in a desired cell or organism. References describing eif6 family members and their biological activities include, for example, the following: Adams et al., Science 87:2185-2195(2000); Wood et al., J. Biol. Chem. 274:11653-11659(1999); and Si et al., Mol. Cell. Biol. 19:1416-1426(1999).
ER	PDOC00992	Enhancer of rudimentary signature	The Drosophila protein 'enhancer of rudimentary' (gene (e(r)) is a small protein of 104 residues whose function is not yet clear. From an evolutionary point of view, it is highly conserved [1] and has been found to exist in probably all multicellular eukaryotic organisms. It has been proposed that this protein plays a role in the cell cycle. As as signaure pattern, we selected a conserved region in the central part of



		}	383
Pfam	Prosite	Full Name	Description
			the protein.
			Description of pattern(s) and/or profile(s)
			Consensus pattern Y-D-I-[SA]-x-L-[FY]-x-F-[IV]-D-x(3)-D-[LIV]-S Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update
			November 1997 / First entry. References
			[1] Gelsthorpe M., Pulumati M., McCallum C., Dang-Vu K., Tsubota S.I.
ER lumen recept	PDOC00732	ER lumen protein	Gene 186:189-195(1997). Proteins that reside in the lumen of the endoplasmic reticulum
a.riamen_recept		retaining receptor signatures	(ER) contain a C-terminal tetrapeptide (generally K-D-E-L or H-D-E-L) that
			serves as a signal for their retrieval (retrograde transport) from subsequent
			compartments of the secretory pathway. The signal is recognized by a receptor
			molecule that is believed to cycle between the cis side of the Golgi apparatus and the ER [1].
	:		This protein is known as the ER lumen protein retaining receptor lor also as
			the 'KDEL receptor'. It has been characterized in a variety of species,
			including fungi (gene ERD2), plants, Plasmodium, Drosophila and mammals. In
			mammals two highly related forms of the receptor are known.
			Structurally, the receptor is a protein of about 220 residues that seems to contain seven transmembrane regions [2]. The N-terminal part (3
			residues) is oriented toward the lumen while the C-terminal tail (about 12
			residues) is cytoplasmic. There are three lumenal and three cytoplasmic loops.
			We developed two signature patterns for these receptors. The first pattern
			as well as
			most of the second transmembrane domain. The second pattern is a perfectly conserved decapeptide that corresponds to the central part of the fifth
			transmembrane domain.
			Description of pattern(s) and/or profile(s)
			Consensus pattern G-[LIV]-S-x-[KR]-x-[QH]-x-L-[FY]-x-[LIV](2)- [FYW]-x(2)-R- Y
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern L-E-[SA]-V-A-I-[LM]-P-Q-[LI] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update
			December 1999 / Patterns and text revised. References
			[1] Pelham H.R.B.



		•	384
Pfam	Prosite	Full Name	Description Curr. Opin. Cell Biol. 3:585-591(1991). [2] Townsley F.M., Wilson D.W., Pelham H.R.B. EMBO J. 12:2821-2829(1993).
ETF_alpha	PDOC00583	Electron transfer flavoprotein alphasubunit signature	The electron transfer flavoprotein (ETF) [1,2] serves as a specific electron acceptor for various mitochondrial dehydrogenases. ETF transfers electrons to the main respiratory chain via ETF-ubiquinone oxidoreductase. ETF is an heterodimer that consist of an alpha and a beta subunit and which bind one molecule of FAD per dimer. A similar system also exists in some bacteria. The alpha subunit of ETF is a protein of about 32 Kd which is structurally related to the bacterial nitrogen fixation protein fixB which could play a role in a redox process and feed electrons to ferredoxin. Other related proteins are: - Escherichia coli hypothetical protein ydiR Escherichia coli hypothetical protein ygcQ. As a signature pattern for these proteins we selected a highly conserved region which is located in the C-terminal section. Description of pattern(s) and/or profile(s) Consensus pattern [LI]-Y-[LIVM]-[AT]-x-G-[IV]-[SD]-G-x-[IV]-Q-H-x(2)-G-x(6)- [IV]-x-A-[IV]-N Sequences known to belong to this class detected by the pattern ALL, except for ygcQ. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / Text revised. References [1] Finocchiaro G., Ikeda Y., Ito M., Tanaka K. Prog. Clin. Biol. Res. 321:637-652(1990). [2] Tsai M.H., Saier M.H. Jr. Res. Microbiol. 146:397-404(1995).
Euk_porin	PDOC00483	Eukaryotic mitochondrial porin signature	The major protein of the outer mitochondrial membrane of eukaryotes is a porin that forms a voltage-dependent anion-selective channel (VDAC) that behaves as a general diffusion pore for small hydrophilic molecules [1 to 4]. The channel adopts an open conformation at low or zero membrane potential and a closed conformation at potentials above 30-40 mV. This protein contains about 280 amino acids and its sequence is composed of between 12 to 16 beta-strands that span the mitochondrial outer membrane. Yeast contains two members of this family (genes POR1 and POR2); vertebrates have at least three members (genes VDAC1, VDAC2 and VDAC3) [5]. As a signature pattern we selected a conserved region located at the C-terminal part of these proteins.

		885
Pfam Prosite	Full Name	Description
F_bP_aidolase PDOC00523	Fructose-bisphosphate aldolase class-II signatures	Description Description of pattern(s) and/or profile(s) Consensus pattern [YH]-x(2)-D-[SPCAD]-x-[STA]-x(3)-[TAG]-[KR]-[LIVMF]- [DNSTA]-IDNS]-x(4)-[GSTAN]-[LIVMA]-x-[LIVMY] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Benz R. Biochim. Biophys. Acta 1197:167-196(1994). [2] Manella C.A. Trends Biochem. Sci. 17:315-320(1992). [3] Dihanich M. Experientia 46:146-153(1990). [4] Forte M., Guy H.R., Mannella C.A. J. Bioenerg. Biomembr. 19:341-350(1987). [5] Sampson M.J., Lovell R.S., Davison D.B., Craigen W.J. Genomics 36:192-196(1996). Fructose-bisphosphate aldolase (EC 4.1.2.13) [1,2] is a glycolytic enzyme that catalyzes the reversible aldol cleavage or condensation of fructose-1,6- bisphosphate into dihydroxyacetone-phosphate and glyceraldehyde 3-phosphate. There are two classes of fructose-bisphosphate aldolases with different catalytic mechanisms. Class-II aldolases [2], mainly found in prokaryotes and fungi, are homodimeric enzymes which require a divalent metal ion generally zinc - for their activity. This family also includes the following proteins: - Escherichia coli galactitol operon protein gatY which catalyzes the transformation of tagatose 1,6-bisphosphate into glycerone phosphate and D-glyceraldehyde 3-phosphate. - Escherichia coli N-acetyl galactosamine operon protein agaY which catalyzes the transformation of tagatose 1,6-bisphosphate into glycerone phosphate and D-glyceraldehyde 3-phosphate. - Escherichia coli N-acetyl galactosamine operon protein agaY which catalyzes the transformation of tagatose 1,6-bisphosphate into glycerone phosphate and D-glyceraldehyde 3-phosphate. - Escherichia coli N-acetyl galactosamine operon protein agaY which catalyzes the same reaction as that of gatY. As signature patterns for this class of enzyme, we selected two conserved regions. The first pattern is located in the first half of the sequence and contains
		Consensus pattern [FYVMT]-x(1,3)-[LIVMH]-[APNT]-[LIVM]-x(1,2)-[LIVM]-H-x-D- H-[GACH] [The two H's are zinc ligands]



		000
Pfam Prosite	Full Name	Description Sequences known to belong to this class detected by the pattern ALL, except for Mycoplasma pneumoniae aldolase. Other sequence(s) detected in SWISS-PROT NONE.
		Consensus pattern [LIVM]-E-x-E-[LIVM]-G-x(2)-[GM]-[GSTA]-x-E Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
		Last update December 1999 / Pattern and text revised.
		References
		[1] Perham R.N. Biochem. Soc. Trans. 18:185-187(1990).
		[2] Marsh J.J., Lebherz H.G. Trends Biochem. Sci. 17:110-113(1992).
		[3] von der Osten C.H., Barbas C.F. III, Wong CH., Sinskey A.J. Mol. Microbiol. 3:1625-1637(1989).
		[4] Berry A., Marshall K.E. FEBS Lett. 318:11-16(1993).
FAA_hydrolase	Fumarylacetoacetate (FAA) hydrolase family	Accession number: PF01557 Definition: Fumarylacetoacetate (FAA) hydrolase family Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_641 (release 4.0) Gathering cutoffs: 25 25 Trusted cutoffs: 42.10 42.10 Noise cutoffs: -93.10 -93.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97255958 Reference Title: Mutations in the fumarylacetoacetate hydrolase gene causing Reference Title: hereditary tyrosinemia type I: overview. Reference Author: St-Louis M, Tanguay RM; Reference Number: [2] Reference Number: [2] Reference Medline: 96125235 Reference Title: Molecular characterization of the 4-hydroxyphenylacetate Reference Title: catabolic pathway of Escherichia coli W: engineering a
		Reference Title: mobile aromatic degradative cluster. Reference Author: Prieto MA, Diaz E, Garcia JL; Reference Location: J Bacteriol 1996;178:111-120. Reference Number: [3] Reference Medline: 96016123 Reference Title: Fungal metabolic model for human type I hereditary Reference Title: tyrosinaemia.
		Reference Author: Fernandez-Canon JM, Penalva MA; Proc Natl Acad Sci U S A 1995;92:9132-9136. Reference Number: [4] 94039092
	·	Reference Title: Purification, nucleotide sequence and some properties of a Reference Title: bifunctional isomerase/decarboxylase from the
		Reference Title: homoprotocatechuate degradative pathway of Escherichia coli Reference Title: C.
		Reference Author: Roper DI, Cooper RA; Reference Location: Eur J Biochem 1993;217:575-580. Database reference: MIM; 276700; Database Reference INTERPRO; IPR002529;



		8	387
Pfam	Prosite	Full Name	Description This family consists of furnishing systems of the system of the systems of the systems of the systems of the syste
			Comment: This family consists of fumarylacetoacetate
			(FAA) hydrolase, Comment: or fumarylacetoacetate hydrolase (FAH) and
			it also includes
			Comment: HHDD isomerase/OPET decarboxylase
			from E. coli strain W.
			Comment: FAA is the last enzyme in the tyrosine catabolic pathway, it hydrolyses
			Comment: fumarylacetoacetate into fumarate and
			acetoacetate which then join the
			Comment: citric acid cycle [1]. Mutations in FAA cause
			type I tyrosinemia in humans
			Comment: this is an inherited disorder mainly affecting the liver leading to
			Comment: liver cirrhosis, hetpatocellular carcinoma,
			renal tubular damages and
			Comment: neurologic crises amongst other symptoms
			[1]. The enzymatic defect causes
			Comment: the toxic accumulation of phenylalanine/tyrosine catabolites [3].
			Comment: The E. coli W enzyme HHDD
			isomerase/OPET decarboxylase contains two
	1		Comment: copies of this domain and functions in fourth
			and fifth steps of the
			Comment: homoprotocatechuate pathway; Comment: here it decarboxylates OPET to HHDD and
			Comment: here it decarboxylates OPE i to HHDD and isomerizes this to OHED.
			Comment: The final products of this pathway are
			pyruvic acid and succinic
			Comment: semialdehyde.
			Number of members: 33
CAD hinding		FAD binding domain	Accession number: PF00667
FAD_binding		FAD bitiding domain	Definition: FAD binding domain
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_180 (release 2.1) Gathering cutoffs: 16.8 16.8
			Gathering cutoffs: 16.8 16.8 Trusted cutoffs: 24.60 16.80
			Noise cutoffs: 13.50 15.90
			HMM build command line: hmmbuild -f HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1] Reference Medline: 95386502
			Reference Title: 53360002 The flavin reductase activity of the
			flavoprotein component
			Reference Title: of sulfite reductase from Escherichia coli. A
			new model for
			Reference Title: the protein structure. Reference Author: Eschenbrenner M, Coves J, Fontecave M
			Reference Location: J Biol Chem 1995;270:20550-20555.
			Reference Number: [2]
			Reference Medline: 96049560
			Reference Title: NADPH-sulfite reductase flavoprotein from
			Escherichia coli: Reference Title: contribution to the flavin content and
	ľ	İ	subunit interaction.
			Reference Author: Eschenbrenner M, Coves J, Fontecave M
			Reference Location: FEBS Lett 1995;374:82-84.
			Reference Number: [3] Reference Medline: 94360001
			Reference Medline: 94360001 Reference Title: Dissection of NADPH-cytochrome P450
			oxidoreductase into
			Reference Title: distinct functional domains.
			Reference Author: Smith GC, Tew DG, Wolf CR;
			Reference Location: Proc Natl Acad Sci U S A 1994;91:8710-
			8714.
			Reference Number: [4] Reference Medline: 97385116
			Reference Title: Three-dimensional structure of NADPH-
			cytochrome P450
			Reference Title: reductase: prototype for FMN- and FAD-
			containing enzymes. Deference Author: Wang M Roberts DI Paschke B Shea
			Reference Author: Wang M, Roberts DL, Paschke R, Shea

		888
rfam Prosi	e Full Name	Description TM Masters BS Kim III
		TM, Masters BS, Kim JJ; Reference Location: Proc Natl Acad Sci U S A 1997;94:8411 8416.
		Database Reference: SCOP; 1amo; fa; [SCOP-USA][CATH-PDBSUM]
		Database Reference INTERPRO; IPR001709;
		Database Reference PDB; 1amo A; 274; 493;
		Database Reference PDB; 1amo B; 274; 493;
		Database Reference PDB; 1quf; 77; 120;
		Database reference: PFAMB; PB001390;
		Comment: This domain is found in sulfite reductase, NADPH cytochrome P450
		Comment: reductase and Nitric oxide synthase. Number of members: 87
AD_binding_3	FAD binding domain	Accession number: PF01494
		Definition: FAD binding domain Author: Bashton M, Bateman A
		Alignment method of seed: Clustalw
	į	Source of seed members: Pfam-B 549 (release 4.0)
		Gathering cutoffs: -7 -7
		Trusted cutoffs: -6.20 -6.20
		Noise cutoffs: -7.90 -7.90
		HMM build command line: hmmbuild -F HMM SEED
		HMM build command line: hmmcalibrateseed 0 HMM
		Reference Number: [1]
İ		Reference Medline: 93028353
		Reference Title: Crystal structure of the reduced form of p-
	i	hydroxybenzoate Reference Title: hydroxylase refined at 2.3A resolution.
		Reference Author: Schreuder HA, van der Laan JM, Swarte
		MB, Kaik KH, Hol WG,
		Reference Author: Drenth J;
		Reference Location: Proteins 1992;14:178-190.
		Database Reference: SCOP; 2phh; fa; [SCOP-USA][CATH-PDBSUM]
		Database Reference INTERPRO; IPR002938;
		Database Reference PDB; 1pxa; 5; 35;
		Database Reference PDB; 1bf3; 5; 139;
		Database Reference PDB; 1bgj; 5; 139;
		Database Reference PDB; 1bgn; 5; 139;
		Database Reference PDB; 1bkw; 5; 139;
		Database Reference PDB; 1cc4 A; 5; 139;
		Database Reference PDB; 1cc6 A; 5; 139;
		Database Reference PDB; 1cj2 A; 5; 139;
		Database Reference PDB; 1pbb; 5; 139; Database Reference PDB; 1pbc; 5; 139;
		Database Reference PDB; 1pbd; 5; 139;
		Database Reference PDB; 1pbe; 5; 139;
		Database Reference PDB; 1pbf; 5; 139;
ı		Database Reference PDB; 1pdh; 5; 139;
		Database Reference PDB; 2phh; 5; 139;
		Database Reference PDB; 1cj3 A; 5; 139;
1		Database Reference PDB; 1cj4 A; 5; 139;
		Database Reference PDB; 1phh; 5; 139;
		Database Reference PDB; 1d7l A; 5; 139;
		Database Reference PDB; 1dob; 5; 139; Database Reference PDB; 1doc; 5; 139;
		Database Reference PDB; 1dod; 5; 139;
		Database Reference PDB; 1doe; 5; 139;
		Database Reference PDB; 1ius; 5; 139;
		Database Reference PDB; 1iut; 5; 139;
		Database Reference PDB; 1iuu; 5; 139;
		Database Reference PDB; 1iuv; 5; 139;
		Database Reference PDB; 1iuw; 5; 139;
		Database Reference PDB; 1iux ; 5; 139;
	İ	Database Reference PDB; 1foh A; 10; 151;
		Database Reference PDB; 1foh D; 10; 151;
1		Database Reference PDB; 1foh B; 10; 151;
		Database Reference PDB; 1foh C; 10; 151;
		Database reference: PFAMB; PB040546;
		Comment: This domain is involved in FAD binding in
1	1	number of enzymes. Number of members: 52





			89
Pfam	Prosite	Full Name	Description
FAD_binding_4	PDOC00674	Oxygen oxidoreductases covalent FAD-binding site	Some oxygen-dependent oxidoreductases are flavoproteins that contains a covalently bound FAD group which is attached to a histidine via an 8-alpha- (N3-histidyl)-riboflavin linkage. These proteins are:
			- 6-hydroxy-D-nicotine oxidase (EC 1.5.3.6) (6-HDNO) [1], a bacterial enzyme that catalyzes the oxygen-dependent degradation of 6-hydroxypicotine into 6-hydroxypyrid-N-methylosmine - Plant reticuline oxidase (EC 1.5.3.9) [2] (berberine-bridge-forming enzyme), an enzyme that catalyzes the oxidation of (S)-reticuline into (S)-scoulerine in the pathway leading to benzophenanthridine alkaloids L-gulonolactone oxidase (EC 1.1.3.8) (I-gulono-gamma-lactone oxidase) [3], a mammalian enzyme which catalyzes the oxidation of L-gulono-1,4-lactone to L-xylo-hexulonolactone which spontaneously isomerizes to L-ascorbate D-arabinono-1,4-lactone oxidase (EC 1.1.3.24) (L-galactonolactone oxidase), a yeast enzyme involved in the biosynthesis of D-erythroascorbic acid [4] Mitomycin radical oxidase [5], a bacterial protein involved in mitomycin resistance and that probably oxidizes the reduced form of mitomycins Rhodococcus fascians fasciation locus protein fas5. The region around the histidine that binds the FAD group is conserved in these enzymes and can be used as a signature pattern.
			Description of pattern(s) and/or profile(s) Consensus pattern P-x(10)-[DE]-[LIVM]-x(3)-[LIVM]-x(9)-[LIVM]-x(3)-[GSA]- [GST]-G-H [H is the FAD binding site] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Text revised. EMBL/GenBank: U40390. References [1] Brandsch R., Hinkkanen A.E., Mauch L., Nagursky H., Decker K. Eur. J. Biochem. 167:315-320(1987). [2] Dittrich H., Kutchan T.M. Proc. Natl. Acad. Sci. U.S.A. 88:9969-9973(1991). [3] Koshizaka T., Nishikimi M., Ozawa T., Yagi K. J. Biol. Chem. 263:1619-1621(1988). [4] Huh WK., Kim ST., Kim JY., Hwang SW., Kang SO.
fer2	PDOC00175; PDOC00642	2Fe-2S ferredoxins, iron- sulfur binding region signature; Adrenodoxin family, iron-sulfur binding	[5] August P.R., Flickinger M.C., Sherman D.H. J. Bacteriol. 176:4448-4454(1994). Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided

	100000000000000000000000000000000000000		390
Pfam	Prosite	Full Name	Description
		region signature	into several subgroups depending upon the physiological nature of the iron
			sulfur cluster(s) and according to sequence similarities. One
			of these
;			subgroups are the 2Fe-2S ferredoxins, which are proteins or
			domains of around
			one hundred amino acid residues that bind a single 2Fe-2S iron- sulfur cluster.
			The proteins that are known [2] to belong to this family are listed
			below.
			Forredovin from photographetic accomismos, nearly plants and
İ			- Ferredoxin from photosynthetic organisms; namely plants and algae where it
			is located in the chloroplast or cyanelle; and cyanobacteria.
			- Ferredoxin from archaebacteria of the Halobacterium genus.
			- Ferredoxin IV (gene pftA) and V (gene fdxD) from Rhodobacter capsulatus.
			- Ferredoxin in the toluene degradation operon (gene xyIT) and
			naphthalene
			degradation operon (gene nahT) of Pseudomonas putida.
			- Hypothetical Escherichia coli protein yfaE.
			- The N-terminal domain of the bifunctional ferredoxin/ferredoxin
			reductase
			electron transfer component of the benzoate 1,2-dioxygenase complex (gene
		į	benC) from Acinetobacter calcoaceticus, the toluene 4-
			monooxygenase complex
			(gene tmoF), the toluate 1,2-dioxygenase system (gene xylZ),
			and the xylene monooxygenase system (gene xylA) from Pseudomonas.
			- The N-terminal domain of phenol hydroxylase protein p5
			(gene dmpP) from
			Pseudomonas Putida.
			- The N-terminal domain of methane monooxygenase component C (gene mmoC)
			from Methylococcus capsulatus .
			- The C-terminal domain of the vanillate degradation pathway
			protein vanB in a Pseudomonas species.
			- The N-terminal domain of bacterial fumarate reductase iron-
			sulfur protein
			(gene frdB).
;			- The N-terminal domain of CDP-6-deoxy-3,4-glucoseen reductase (gene ascD)
			from Yersinia pseudotuberculosis.
			- The central domain of eukaryotic succinate dehydrogenase
			(ubiquinone) iron- sulfur protein.
			- The N-terminal domain of eukaryotic xanthine dehydrogenase.
			- The N-terminal domain of eukaryotic aldehyde oxidase.
			In the 2Fe-2S ferredoxins, four cysteine residues bind the
			iron-sulfur
			cluster. Three of these cysteines are clustered together in the
			same region of
			the protein. Our signature pattern spans that iron-sulfur binding region.
			Description of pattern(s) and/or profile(s)
			=======================================
			Consensus pattern C-{C}-{C}-[GA]-{C}-C-[GAST]-
			{CPDEKRHFYW}-C [The three C's are 2Fe-2S ligands] Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT 15.
			Note in addition to the proteins listed above there are a number of
			other ferredoxin-like proteins that bind a 2Fe-2S cluster but which
			do not seem to be evolutionary related to this family. Among them are the ferredoxins from the adrenodoxin family (see
			<pdoc00642>) as well as the bacterial aromatic dioxygenase</pdoc00642>

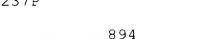


			91
Pfam	Prosite	Full Name	Description Description
			systems ferredoxin-like proteins such as bnzC, ndoA, and todB.
			Last update November 1997 / Text revised.
			References
			[1]
			Meyer J.
			Trends Ecol. Evol. 3:222-226(1988).
			[2] Harayama S., Polissi A., Rekik M.
			FEBS Lett. 285:85-88(1991).
			, ,
			Ferredoxins [1] are a group of iron-sulfur proteins which mediate
			electron
			transfer in a wide variety of metabolic reactions. Ferredoxins can be divided
			into several subgroups depending upon the physiological nature
			of the iron
			sulfur cluster(s) and according to sequence similarities. One
			family of
			ferredoxins groups together the following proteins that all bind a
			single 2Fe- 2S iron-sulfur cluster:
			EG HOLL GRANDI GRANDI
			- Adrenodoxin (ADX) (adrenal ferredoxin), a vertebrate
			mitochondrial protein
			which transfers electrons from adrenodoxin reductase to
			cytochrome P450scc, which is involved in cholesterol side chain cleavage.
			- Putidaredoxin (PTX), a Pseudomonas putida protein which
			transfers electrons
İ			from putidaredoxin reductase to cytochrome P450-cam, which
			is involved in
			the oxidation of camphor Terpredoxin [2], a Pseudomonas protein which transfers
			electrons from
			terpredoxin reductase to cytochrome P450-terp, which is
			involved in the
			oxidation of alpha-terpineol.
			- Rhodocoxin [3], a Rhodococcus protein which transfers
			electrons from rhodocoxin reductase to cytochrome CYP116 (thcB), which is
i			involved in the
			degradation of thiocarbamate herbicides.
			- Escherichia coli ferredoxin (gene fdx) [4] whose exact function
			is not yet
			known Rhodobacter capsulatus ferredoxin VI [5], which may transfer
			electrons to a
			yet uncharacterized oxygenase.
			- Caulobacter crescentus ferredoxin (gene fdxB) [6].
			In these proteins favy avatoins positives hind the iron avitive
			In these proteins, four cysteine residues bind the iron-sulfur cluster. Three
	1		of these cysteines are clustered together in the same region of
			the protein.
			Our signature pattern spans that iron-sulfur binding region.
			Description of pattern(s) and/or profile(s)
			Consensus pattern C-x(2)-[STAQ]-x-[STAMV]-C-[STA]-T-C-[HR]
			[The three C's are 2Fe-2S ligands]
			Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT 1.
]			Last update
			November 1995 / Pattern and text revised.
			EMBL/Genbank: X51607. References
			[1]
			Meyer J. Trends Ecol. Evol. 3:222-226(1988)
			Trends Ecol. Evol. 3:222-226(1988).
	1		

		. 8	392
Pfam	Prosite	Full Name	Description
<u> </u>			[2] Peterson J.A., Lu JY., Geisselsoder J., Graham-Lorence S., Carmona C., Witney F., Lorence M.C. J. Biol. Chem. 267:14193-14203(1992).
			[3] Nagy I., Schoofs G., Compernolle F., Proost P., Vanderleyden J., De Mot R. J. Bacteriol. 177:676-687(1995).
			[4] Ta D.T., Vickery L.E. J. Biol. Chem. 267:11120-11125(1992).
			[5] Naud I., Vincon M., Garin J., Gaillard J., Forest E., Jouanneau Y. Eur. J. Biochem. 222:933-939(1994).
i			[6] Amemiya K
Ferric_reduct		Ferric reductase like	Accession number: PF01794
		transmembrane component	Definition: Ferric reductase like transmembrane component
		Component	Author: Bashton M, Bateman A
			Alignment method of seed: T_Coffee
			Source of seed members: Pfam-B_728 (release 4.2) Gathering cutoffs: -122 -122
			Trusted cutoffs: -34.80 -34.80
			Noise cutoffs: -210.30 -210.30
			HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 93309468
			Reference Title: The fission yeast ferric reductase gene
			frp1+ is required Reference Title: for ferric iron uptake and encodes a protein
			that is Reference Title: homologous to the gp91-phox subunit of
	-		the human NADPH Reference Title: phagocyte oxidoreductase.
			Reference Author: Roman DG, Dancis A, Anderson GJ,
t gés			Klausner RD; Reference Location: Mol Cell Biol 1993;13:4342-4350.
			Reference Number: [2]
			Reference Medline: 92294876
			Reference Title: Cytochrome b558: the flavin-binding component of the
			Reference Title: phagocyte NADPH oxidase.
			Reference Author: Rotrosen D, Yeung CL, Leto TL, Malech
			HL, Kwong CH; Reference Location: Science 1992;256:1459-1462.
			Reference Number: [3]
			Reference Medline: 87258189 Reference Title: The glycoprotein encoded by the X-linked
			Reference Title: The glycoprotein encoded by the X-linked chronic
			Reference Title: granulomatous disease locus is a
			component of the
			Reference Title: neutrophil cytochrome b complex. Reference Author: Dinauer MC, Orkin SH, Brown R, Jesaitis
			AJ, Parkos CA;
			Reference Location: Nature 1987;327:717-720. Reference Number: [4]
,			Reference Number: [4] Reference Medline: 87258190
			Reference Title: The X-linked chronic granulomatous
			disease gene codes for Reference Title: the beta- chain of cytochrome b-245.
			Reference Author: Teahan C, Rowe P, Parker P, Totty N,
			Segal AW;
			Reference Location: Nature 1987;327:720-721. Database Reference INTERPRO; IPR002916;
			Comment: This family includes a common region in the
			transmembrane proteins
			Comment: mammalian cytochrome B-245 heavy chain



nv.			893
Pfam	Prosite	Full Name	Description
			(gp91-phox), ferric reductase
			Comment: transmembrane component in yeast and
			respiratory burst oxidase from Comment: mouse-ear cress.
			Comment: This may be a family of flavocytochromes
			capable of moving electrons
1			Comment: across the plasma membrane [1].
İ			Comment: The Frp1 protein Swiss:Q04800 from S.
			pombe is a ferric reductase
	1		Comment: component and is required for cell surface
		İ	ferric reductase activity, Comment: mutants in frp1 are deficient in ferric iron
l			Comment: mutants in frp1 are deficient in ferric iron uptake [1].
			Comment: Cytochrome B-245 heavy chain
			Swiss:P04839 is a FAD-dependent
			Comment: dehydrogenase it is also has electron
			transferase activity which reduces
			Comment: molecular oxygen to superoxide anion, a
			precursor in the production of Comment: microbicidal oxidants [2]
			Comment: microbicidal oxidants [2]. Comment: Mutations in the sequence of cytochrome B-
			245 heavy chain (gp91-phox)
			Comment: lead to the X-linked chronic granulomatous
			disease. The bacteriocidal
			Comment: ability of phagocytic cells is reduced and is
			characterised by the
			Comment: absence of a functional plasma membrane associated NADPH oxidase [3].
			Comment: The chronic granulomatous disease gene
			codes for the beta chain of
			Comment: cytochrome B-245 and cytochrome B-245 is
			missing from patients with
			Comment: the disease [4].
			Comment: The aligned region includes a potential FAD
			binding domain. Number of members: 34
		! !	Number of members. 34
Flavi_NS5		Flavivirus RNA-directed	Accession number: PF00972
		RNA polymerase	Definition: Flavivirus RNA-directed RNA polymerase
			Author: Finn RD, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_200 (release 3.0) Gathering cutoffs: 12 12
			Trusted cutoffs: 16.00 16.00
	İ		Noise cutoffs: 8.50 8.50
	1		11035 CULOTIS. 0.30 6.30
			HMM build command line: hmmbuild -f HMM SEED
ľ			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus:
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus.
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Reference Author: EA; Reference Location: Virology 1995;206:1133-1139.
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Reference Author: EA; Reference Location: Virology 1995;206:1133-1139. Reference Number: [2]
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Reference Author: EA; Reference Location: Reference Number: [2] Reference Medline: 96182933
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: 95159427 Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Reference Number: [2] Reference Medline: 96182933 Recombinant dengue type 1 virus NS5
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Reference Number: [2] Reference Medline: 96182933 Recombinant dengue type 1 virus NS5
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: 95159427 Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Virology 1995;206:1133-1139. [2] Reference Medline: P6182933 Recombinant dengue type 1 virus NS5 Reference Title: Escherichia coli exhibits RNA-dependent
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Reference Number: Reference Medline: Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Number: [2] Reference Medline: Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Number: Reference Number: [2] Reference Title: Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Number: Reference Number: Reference Number: Reference Title: Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Number: Reference Number: Reference Number: Reference Number: Reference Number: Reference Title: Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Number: Reference Number
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Reference Number: [2] Reference Medline: 96182933 Recombinant dengue type 1 virus NS5 Protein expressed in Reference Title: protein expressed in Reference Title: RNA polymerase Reference Title: activity. Reference Author: Tan BH, Fu J, Sugrue RJ, Yap EH, Chan
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: Reference Medline: 95159427 Reference Title: evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Reference Number: Reference Medline: Reference Medline: Reference Title: protein expressed in Reference Title: protein expressed in Reference Title: RNA polymerase Reference Title: Reference Author: YC, Tan YH;
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Reference Number: [2] Reference Medline: 96182933 Reference Title: Potein expressed in Reference Title: Protein expressed in Reference Title: RNA polymerase Reference Title: RNA polymerase Reference Title: RNA polymerase Reference Title: RNA polymerase Reference Title: RNA polymerase Reference Title: RNA polymerase Reference Title: RNA polymerase Reference Author: YC, Tan YH; Reference Location: Virology 1996;216:317-325.
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: 95159427 Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Virology 1995;206:1133-1139. [2] 96182933 Recombinant dengue type 1 virus NS5 Protein expressed in Reference Title: Escherichia coli exhibits RNA-dependent RNA polymerase Reference Author: YC, Tan YH; Reference Location: Reference Location: Reference Number: [3]
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Wirology 1995;206:1133-1139. [2] 96182933 Recombinant dengue type 1 virus NS5 Protein expressed in Reference Title: RNA polymerase Reference Title: RNA polymerase Reference Title: Roference Author: YC, Tan YH; Reference Location: Reference Location: Reference Number: Reference Medline: Possible for the more alignment of the
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Author: Phylogeny of TYU, SRE, and CFA virus: difference Author: Phylogeny of TYU, SRE, and CFA viru
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Author: Phylogeny of TYU, SRE, and CFA virus: difference File: Phylogeny of TYU, SRE, and CFA virus:
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] P5159427 Reference Medline: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Author: Phylogeny of TYU, SRE, and CFA virus: different Reference Author: Phylogeny of TYU, SRE, and CFA virus: different Reference Author: Phylogeny of TYU, SRE, and CFA virus: different Phylogeny of TYU, SRE, and CFA virus:
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: Reference Medline: 95159427 Reference Title: different Reference Title: Reference Author: EA; Reference Location: Reference Number: Reference Medline: Potein expressed in Reference Title: protein expressed in Reference Title: RNA polymerase Reference Title: RNA polymerase Reference Author: YC, Tan YH; Reference Location: Reference Location: Reference Number: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Title: Million Million Reference Medline: Reference Title: Million Reference Medline: Reference Title: Million Reference Medline: Reference Title: Million Reference Medline: Reference Title: Million Reference Medline: Reference Medline: Reference Medline: Reference Medline: Reference Title: Million Reference Medline: Reference Title: Million Reference Medline: Reference Title: Million Reference Medline: Reference Title: Million Reference Medline: Reference Title: Million Reference Medline: Reference Medline: Reference Title: Million Reference Medline: Reference Medline: Reference Author: Yirology 1995;206:1133-1139. Recombinant dengue type 1 virus NS5 Escherichia coli exhibits RNA-dependent Reference Medline: Activity. Tan BH, Fu J, Sugrue RJ, Yap EH, Chan Yirology 1996;216:317-325. [3] 93224895 Computer-assisted identification of a methyltransferase domain in NS5 protein of
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: different Reference Title: Reference Title: Reference Author: EA; Reference Location: Reference Number: Reference Medline: Reference Title: Reference Title: Reference Title: Reference Title: Reference Number: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Title: Reference Author: YC, Tan YH; Reference Location: Reference Number: Reference Medline: Reference Medline: Reference Medline: Reference Title: Reference Number: Reference Author: Ycrology 1995;206:1133-1139. Recombinant dengue type 1 virus NS5 Escherichia coli exhibits RNA-dependent Reference Number: Refer



Fork_head PDOC00564 Port Position Posi			8	394
Reference Mediline: Reference Title: RPMA viruses: Reference Title: Reference Title: Reference Title: Reference Author:	Pfam	Prosite	Full Name	Description
Reference Tille: RNA viruses: Reference Tille: Reference Tille: Reference Tille: Reference Tille: Reference Tille: Reference Author: Reference Author: Reference Author: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Location: Reference Referen				
RNA viruses: Reference Title: Amino acid Reference Title: Reference Alpha- Reference Alpha- Reference Alpha- Reference Alpha- Reference Alpha- Reference Alpha- Reference Alpha- Reference Alpha- Reference Alpha- Reference Alpha- Reference Alpha- Reference Comment: Serioviruse produce a polyprotein from the serioviruse produce a polyprotein from the serioviruse produce a polyprotein from the serioviruse produce a polyprotein from the serioviruse produce a polyprotein from the serioviruse produce a polyprotein from the serioviruse produce a polyprotein from the serioviruse from the serioviruse produce a polyprotein from the serioviruse from the seriovirus produce a polyprotein from the serioviruse produce and profession and motification from the seriovirus produce and profession and seriovirus produce and profession and seriovirus produce and profession and seriovirus produce and profession and seriovirus produce and profession and seriovirus produce and profession and seriovirus produce and profession from the seriovirus produce and profession factor and profession from the seriovirus produce and profession factor and profession factor and profession factor and profession factor and profession factor and profession factor and profession factor and profession factor and profession factor and profession factor and profession factor and profession factor and profession factor (LEP.) LIE buptine-rich NPAT-like motifs in the HIV-1 LTR and the intereutkin-2 promoter. LIE may be invoked in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian interleukin-enhancer binding factor (LEP.) LIE buptine-rich NPAT-like motifs in the HIV-1 LTR and the intereutkin-2 promoter. LIE may be invoked in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor EF-1 which plays an important celabilishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLE, a protein				
Reference Title: amino acid Reference Althor. Reference Althor. Reference Althor. Reference Althor. Database Reference Comment: SRNA genome. Comment: The Additional Polymerase Comment: Comment: The Additional Polymerase Comment: The Additional Polymerase Comment: The Additional Polymerase Comment: The Additional Polymerase [2]. Number of members: Top polymerase [2]. Numbers of members: Top polymerase [2]. Numbers of members: Top polymerase [2]. Numbers of members: Top polymerase [2]. Numbers of members: Top polymerase [
amino acid Reference Author: Reference Author: Reference Author: Reference Author: Reference Cucinion: Comment: SRNA genome. Comment: SRNA genome. Comment: This protein is also known as NS5. This RNA directed RNA polymerase possesses a number of short comment: regions and motifs homologous to other polymerase [2]. Number of members: This protein is also known as NS5. This RNA directed RNA polymerase possesses a number of short comment: Polymerase [2]. Number of members: This protein is also known as NS5. This RNA directed RNA polymerase possesses a number of short comment: polymerases [2]. Number of members: This protein is also known as NS5. This RNA directed RNA polymerase possesses a number of short contain a consared domain of about 100 amino-acid residues, called domain (but also known as a "lunged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophilia protein crocodile (gene croc) [3], which is required for the croth of the				
Reference Author. Reference Author. Reference Author. Reference Author. Reference Author. Reference Author. Reference Author. Reference Author. Reference Author. Reference Author. Reference Coation. Reference Coation. Reference Coation. Reference Coation. Reference Coation. Reference Coation. Reference Coation. Reference Interest Produce a polyprotein from the send of the Coation. Reference Comment. This protein is also known as NSS. Comment. This protein is also known as NSS. Comment. Report Produce a polyprotein from the send of the Reference Coation. Reference Author. Reference				
Reference Author: Reference Author: Reference Coation: Crit Rev Blochem Mol Biol 1993;28:375-430. Database Reference Comment: SeRNA genome. Comment: Flaw/iruses produce a polyprotein from the seRNA genome. Comment: This protein is de lor known as NS5. This RNA-directed RNA polymerase possesses a number of short regions and motifs homologous to other polymerase [2]. Number of members: TSP polymerases [2]. Number of members and protein feet of serious polymerases [2]. Number of members in the serious polymerases [2]. Number of members in the serious polymerases [2]. Number of members in the purpose [2]. Number of members in the purpose [2]. Number of members in the purpose [2]. Number of members in the purpose [2]. Number of members in the purpose [2]. Number of members in the purpose [2]. Number of members in the purpose [2]. Number of members in the purpose [2]. Number of m				
Additional process of the serious field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the field of the f				
Database Reference Comment: Flaviviruse produce a polyprotein from the sRNA genome. Comment: This protein is also known as NS5. Comment: This Provision and motifs homologous to other possesses a number of short in the serious and motifs homologous to other polyprotein signatures and profile in the serious and motifs homologous to other provision and motifs homologous to other provision and motifs homologous to other provision and motifs homologous to other provision and profile in the serious provision and profile occuration a conserved domain of about 100 amino-acid residues, called the fork head or a winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila protein croccodile (gene croc) [3], which is required for the protein protein protein proteins protein proteins prote				Reference Location: Crit Rev Biochem Mol Biol 1993;28:375-
Comment: SaRNA genore: Comment: This protein is also known as N55. Comment: This RNA-directed RNA polymerase possesses a number of short comment: regions and motifs homologous to other RNA-directed RNA polymerases [2]. Number of members: 159 Fork_head PDOC00584 Fork head domain signatures and profile by the second of members: 159 I have been shown [1] that some eukaryotic transcription factors contain a conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in the regulates the expression of genes involved in the regulates the expression of genes involved in a conserved of the establishment of head structures. - Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins stoppy paired 1 and 2 (slip1 and slp2) involved in a conserved in a conserved in the sericin-1 gene. - Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes. - Mammalian transcription factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be travolved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor EFL-1 which plays an important role in the establishment of the regional subdivision of the developing binds in the regional subdivision of the developing binds in the regional subdivision of the developing binds in the regional subdivision of the developing binds in the regional subdivision of the developing binds in the regional subdivision of the developing binds in the regional subdivision of the developing binds in the regional subdivision of the developing binds in the regional subdivision of the developing bin				
ssRNA genome: Comment: This protein is also known as NS5. Comment: This RNA-directed RNA polymerase possesses a number of short RNA-directed RNA Comment: RNA-directed RNA Comment: RNA-directed RNA Comment: RNA-directed RNA Comment: Signatures and profile It has been shown [1] that some eukaryotic transcription factors contain a conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (Rh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal deminance of the separation. - Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericin-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and purise-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian interactivith and cellular promoter elements Mammalian interactivith and cellular promoter elements Mammalian transcription factor BF-1 which plays an important cestabilishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTF-8 protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-1 LTR) Mammalian transcription factors FREAC-1 (FKHL5), FFEAC-5 (FKHL5), FFHA-1, FKH-13), FRHA-3, FRHA-4, FKH-15, FKH-11, FKH-11, FRHA-3, FRHA-6 FREAC-2 (FKH-15), FKH-3, FKH-13, FKH-3, FKH-14, FKH-15, FKH-13, FKH-3, FKH-14, FKH-15, FKH-15, FKH-13, FKH-13, FKH-14, FKH-15, FKH-15, FKH-15, FKH-15, FKH-16, FKH-16, FKH-16, FKH-16, FKH-16, FKH-16, FKH-16, FKH-16, FKH-16, FKH-16, FKH-16, FKH-16, FKH-1				
Comment: This protein is also known as NS5. Comment: This RNA-directed RNA polymerase possesses a number of short comment: regions and motifs homologous to other RNA-directed RNA comment: polymerases [2]. Number of members: 159 Fork_head PDOC00564 Fork head domain signatures and profile the shown [1] that some eukaryotic transcription factors conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (Rkh), Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila proteins rocodile (gene croc) [3], which is required for the establishment of head structures. - Drosophila proteins rocodile (gene croc) [3], which is required for the establishment of head structures. - Drosophila proteins sloppy paired 1 and 2 (sip1 and sip2) involved in a sericim-1 gene. - Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver gene. - Mammalian transcription factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interieukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor EFI-1 which plays an important role in the establishment of the regional subdivision of the developing bits development of the telencephalon. - Human AT-EI. - Human AT-EI. - Haman HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-1 LTR). - Mammalian transcription factors FREA-C1 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (FKHLS), FFIEA-C5 (
Comment: This RNA-directed RNA polymerase possesses a number of short Comment: regions and motifs homologous to other RNA-directed RNA Comment: Polymerase [2]. Number of members: 159 Fork_head PDOC00564 Fork head domain signatures and profile in the seek of the seek				l
Comment: polymerases [2]. Number of members: 159 It has been shown [1] that some eukaryotic transcription factors conserved domain signatures and profile It has been shown [1] that some eukaryotic transcription factors conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (Rkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures. - Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation. - Bombyx mori slik gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene. - Mammalian transcriptional activators HNF-3-alpha, -beta, and gamma. The HT-3 proteins interact with the cis-acting regulatory regions of a liver genes. - Mammalian interact with the cis-acting regulatory regions of a liver genes. - Mammalian interactive with the cis-acting regulatory regions of a liver genes. - Mammalian interactive hother hostis in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia viral and cellular promoter elements. - FREAC-3 (FKH-1, FKH-1), FREAC-4 (FKH-1, FREAC-5 (FKH-1, FKH-2, HFH-3), FKH-3, FKH-3, FKH-3, FKH-3, FKH-4, FKH-3, FKH-3, FKH-4, FKH-5, FKH-3, FKH-4, FKH-1, FKH-3, FKH-3, FKH-4, FKH-1, FKH-3, FKH-3, FKH-4, FKH-3, FKH-4, FKH-3, FKH-4, FKH-3, FKH-3, FKH-4, FKH-3, FKH-4, FKH-3, FKH-4, FKH-3, FKH-4, FKH-3, FKH-4, FKH				
Fork_head PDOC00564 Fork head domain signatures and profile It has been shown [1] that some eukaryotic transcription factors contain a conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (Rh), Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila proteins crocodile (gene croc) [3], which is required for the establishment of head structures Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-graphic proteins and in the HIV-1 LTR and the interleukin-alman transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephaton Human H-IF-, a protein that binds to the purine-rich region in the development of the telencephaton Human H-IF-, a protein that binds to the purine-rich region in the development of the telencephaton Human H-IF-, a protein that binds to the purine-rich region in the development of the telencephaton Human H-IF-, a protein that binds to the purine-rich region in the development of the telencephaton Human H-IF-, a protein that binds to the purine-rich region in the development of the telencephaton Human H-IF-, a protein that binds to the purine-rich region in the development of the telencephaton Human H-IF-, a protein that binds to the purine-rich region in FREAC-3 (FKH-IL), FREAC-4 (FKH-				
Comment: polymerases [2]. Number of members: 15 Number of members: 15 Number of members: 15 Number of members: 15 Number of members: 15 Number of members: 15 Number of members: 15 Number of members: 15 Number of members: 15 Number of members: 15 Number of members: 16 Nume				
PDOC00564 Fork head domain signatures and profile to contain a conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fikh). Fikh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila proteins reD2, FD3, FD4, and FD5 Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins red in the sericin-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The - HHF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian transcriptional activators HNF-1 and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor EF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human H-IF, a protein that binds to the purine-rich region in the development of the telencephalon Human H-IF, a protein that binds to the purine-rich region in the development of the telencephalon Human H-IF, a protein that binds to the purine-rich region in the development of the telencephalon Human H-IF, a protein that binds to the purine-rich region in the development of the telencephalon Human H-IF, a protein that binds to the purine-rich region in the development of the telencephalon Human H-IF, a protein that binds to the purine-rich region in the development of the telencephalon Human H-IF, a protein that binds to the purine-rich region in the development of the telencephalon Human AFIX which is involved in a chromosomal translocation that causes acute l				l
PDOC00564 Fork head domain signatures and profile contain a conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interect with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-enhancer binding factor (ILF) and the interleukin-enhancer binding factor (ILF) and the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important cle in the establishment of the regional subdivision of the developing brain and in the development of the telencephation Human HTLF, a protein that binds to the purine-rich region in human T-cell leukema virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHLS), HFH-8), FREAC-2 (FKHLG), FRH-1), FREAC-3 (FKHLG), FRH-1), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL11), FREAC-8 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FREAC-6 (FKHL10, HFH-4), FRE				
signatures and profile conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures Drosophila proteins FDZ, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori slik gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and gamma. The HHF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binding factors interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukema virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-2 (FKHL6), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-6 (FKHL1), FREAC-6 (FKHL1), FREAC-7 (FKHL1), FREAC-8 (FKHL9), FREAC-6 (FKHL9), FREAC-6 (FKHL9), FREAC-6 (FKHL9), FREAC-6 (FKHL9), FREAC-7 (FKHL1), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FK				Number of members. 159
signatures and profile conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures Drosophila proteins FDZ, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori slik gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and gamma. The HHF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binding factors interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukema virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-2 (FKHL6), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-6 (FKHL1), FREAC-6 (FKHL1), FREAC-7 (FKHL1), FREAC-8 (FKHL9), FREAC-6 (FKHL9), FREAC-6 (FKHL9), FREAC-6 (FKHL9), FREAC-6 (FKHL9), FREAC-7 (FKHL1), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-8 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FKHL9), FREAC-9 (FK	Fork head	PDOC00564	Fork head domain	It has been shown [1] that some eukarvotic transcription factors
the fork head domain (but also known as a "winged helix"), which is involved in DNA-bindring [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in a segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sercim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of ilver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHLS, HFH-8), FREAC-2 (FKHLI), FREAC-3 (FKHLI), FREAC-3 (FKHLI), FREAC-4 (FKHLB), FREAC-6 (FKHLI), FREAC-6 (FKHLI), FREAC-7 (FKHLI), FREAC-8 (FKHLI), FREAC-7 (FKHLI), FREAC-8 (FKHLI), FREAC-7 (FKHLI), FREAC-8 (FKHLI), FREAC-7 (FKHLI), FREAC-8 (FKHLI), FREAC-8 (FKHLI), FREAC-8 (FKHLI), FREAC-8 (FKHLI), FREAC-6 (FKHLI), FREAC-8 (FKHLI), FREAC-6 (FKHLI), FREAC-6 (FKHLI), FREAC-7 (FKHLI), FREAC-8 (FKHLI), FREAC-8 (FKHLI), FREAC-8 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FREAC-9 (FKHLI), FR	T OIK_IICAG	1 20000001		
domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHLS, HFH-8), FREAC-2 (FKHLS), FREAC-3 (FKHLT), FREAC-4 (FKHLS), FREAC-5 (FKHLS), FREAC-6 (FKHLS), FREAC-7 (FKHLS), FREAC-6 (FKHLS), FREAC-7 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-7 (FKHLS), FREAC-8 (FKHLS), FREAC-7 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-6 (FKHLS), FREAC-8 (FKHLS), FREAC-6 (FKHLS), FREAC-7 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREA				conserved domain of about 100 amino-acid residues, called
DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures. - Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of transcription of transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of ilver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purifier-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHLS, HFH-8), FREAC-2 (FKHLS), FREAC-3 (FKHLT), FREAC-4 (FKHLB), FREAC-5 (FKHLS), FREAC-5 (FKHLT), FREAC-6 (FKHLS), FREAC-6 (FKHLS), FREAC-7 (FKHLS), FREAC-6 (FKHLS), FREAC-7 (FKHLS), FREAC-8 (FKHLS), FREAC-7 (FKHLS), FREAC-8 (FKHLS), FREAC-6 (FKHLS), FREAC-7 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-6 (FKHLS), FREAC-6 (FKHLS), FREAC-6 (FKHLS), FREAC-7 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-8 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC-9 (FKHLS), FREAC				
[2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures. - Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins Sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation. - Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian reascription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-3 (FKHL5, FKH-1, FKH-5, FREAC-4 (FKHL10), FREAC-6 (FKHL10), FFREAC-6 (FKHL10), FFREAC-6 (FKHL10), FFREAC-6 (FKHL10), FFREAC-7 (FKHL11), FREAC-8 (FKHL1), FFREAC-6 (FKHL10), FFREAC-6 (FKHL10), FFREAC-7 (FKHL11), FFREAC-8 (FKHL1), FFREAC-8 (FKHL1), FFREAC-8 (FKHL10), FFREAC-6 (FKHL10), FFREAC-7 (FKHL11), FFREAC-8 (FKHL10), FFREAC-6 (FKHL10), FFREAC-6 (FKHL10), FFREAC-7 (FKHL111), FFREAC-8 (FKHL10), FFREAC-6 (FKHL10), FFREAC-7 (FKHL111), FFREAC-8 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-9 (FKHL10), FFREAC-				
- Drosophila fork head protein (fkh), Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures. - Drosophila proteins FD2, FD3, FD4, and FD5. - Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation. - Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene. - Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes. - Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5), FREAC-3 (FKHL5), FREAC-3 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-6), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-6), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia acute.		}		
transcription factor that regulates the expression of genes involved in terminal development Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-3 (FKHL12, FKH-1), FREAC-4 (FKHL13, FREAC-5 (FKHL13, FKH-1), FREAC-4 (FKHL13, FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL11, FRH-1, FKH-5), FREAC-7 (FKHL11), FREAC-8 (FKH-1, FKH-6, HFH-1, HH-15), FREAC-7 (FKHL11), FREAC-8 (FKH-1, FKH-6, HFH-1, HH-15), FREAC-7 (FKHL11), FREAC-8 (FKH-1, FKH-6, HFH-1, HH-15), FREAC-7 (FKHL11), FREAC-8 (FKH-1, FKH-6, HFH-6), FKH-6, HFH-1, FKH-6, HFKH-1, FKH-6, HFH-1, FKH-6, HFH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, F				[2]. Proteins known to contain this domain are noted below.
transcription factor that regulates the expression of genes involved in terminal development Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-3 (FKHL12, FKH-1), FREAC-4 (FKHL13, FREAC-5 (FKHL13, FKH-1), FREAC-4 (FKHL13, FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL11, FRH-1, FKH-5), FREAC-7 (FKHL11), FREAC-8 (FKH-1, FKH-6, HFH-1, HH-15), FREAC-7 (FKHL11), FREAC-8 (FKH-1, FKH-6, HFH-1, HH-15), FREAC-7 (FKHL11), FREAC-8 (FKH-1, FKH-6, HFH-1, HH-15), FREAC-7 (FKHL11), FREAC-8 (FKH-1, FKH-6, HFH-6), FKH-6, HFH-1, FKH-6, HFKH-1, FKH-6, HFH-1, FKH-6, HFH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, FKH-6, HFKH-1, F				- Drosophila fork head protein (fkh). Fkh is probably a
development. - Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures. - Drosophila proteins FD2, FD3, FD4, and FD5. - Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation. - Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene. - Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes. - Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTV-1 LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-2 (FKHL6), FREAC-3 (FKHL9, FKH-1), FREAC-4 (FKHL11), FREAC-8 (FKHL12, HFH-4), FKH-3, FFH-4. A Human AFA1 which is involved in a chromosomal translocation that causes acute leukemia.				transcription factor
- Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL10, HFH-5), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
for the establishment of head structures. Drosophila proteins FD2, FD3, FD4, and FD5. Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation. Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene. Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes. Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFA1-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, HFH-6), FREAC-4 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. Human AFX1 which is involved in a chromosomal transiocation that causes acute leukemia.				
establishment of head structures. - Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell letwem virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFXI which is involved in a chromosomal translocation that causes acute leukemia.				
- Drosophila proteins FD2, FD3, FD4, and FD5 Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, HFH-6), FREAC-6 (FKHL10, HFH-1), FRH-3, FKH-3, HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
- Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FKH-2, HFH-6), FREAC-6 (FKHL12, HFH-7), FKH-3, HFH-6), FREAC-6 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, HFH-4, H-4, H-H-4, H				
segmentation. - Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene. - Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes. - Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL1), FREAC-5 (FKHL1, FKH-1, FRH-1), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL10, HFH-5), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-6), FREAC-6 (FKHL10, HFH-7), FREAC-8 (FKHL11, HFH-7), FREAC-8 (FKHL12, HFH-6), FREAC-6 (FKHL10, HFH-7), FREAC-8 (FKHL10, HFH-7), FREAC-9 (FKHL11), FREAC-8 (FKHL10, HFH-7), FREAC-9 (FKHL10, HFH-7), FREAC-9 (FKHL11), FREAC-8 (FKHL10, HFH-7), FREAC-9 (FKHL10), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), FREAC-9 (FKHL10), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), FREAC-9 (FKHL10), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10), HFH-7), FREAC-9 (FKHL10),				- Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2)
- Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL6), FREAC-4 (FKHL8), FREAC-5 (FKHL12, HFH-7), FKH-3), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FRH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
transcription of the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL12, HFH-6), FREAC-6 (FKHL10, HFH-7), FKH-3, FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FFREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-1.				segmentation.
the sericim-1 gene Mammalian transcriptional activators HNF-3-alpha, -beta, and -garma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-4, FKH-1), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
- Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes. - Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephaton. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FRH-6, FRH-C-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes. - Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FRHAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
a number of liver genes. - Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHLS, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL10, HFH-5), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				-gamma. The
of liver genes. - Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-2 (FKHL16), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-7), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FREAC-8 (FKHL13, HFH-7), FKH-3, FREAC-9 (FKHL14, HFH-7), FREAC-9 (FKHL14, HFH-7), FREAC-9 (FKHL14, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
- Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6). FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FRH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-3 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.			ľ	live de la disc
interleukin-2 promoter: ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.	1			purine-rich NFAT-like motifs in the HIV-1 LTR and the
regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				interleukin-2
important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				,
- Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.		1		·
brain and in the development of the telencephalon Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.		1		
- Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				brain and in
human T-cell leukemia virus long terminal repeat (HTLV-I LTR) Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
teukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.		1		
- Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL6), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				FREAC-2 (FKHL6),
(FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5
(FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				(FKHL9, FKH-2, HFH-6),
FKH-4, FKH-5, HFH-1 and HFH-4 Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
- Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia.				
translocation that causes acute leukemia.				- Human AFX1 which is involved in a chromosomal
acute leukemia.				
- Human FKHR which is involved in a chromosomal		1		
				- Human FKHR which is involved in a chromosomal



	1_		895
Pfam	Prosite	Full Name	Description
			translocation that causes
			rhabdomyosarcoma.
			- Xenopus XFKH1, a protein essential for normal axis formation Caenorhabditis elegans lin-31; involved in the regulation of
			- Caenornabditis elegans lin-31; involved in the regulation of vulval cell
			fates.
			- Yeast HCM1, a protein of unknown function.
			- Yeast FKH1.
			- Yeast FKH2.
			The fork domain is highly conserved. We have developed two
			patterns for its
			detection. The first corresponds to the N-terminal section of the
			domain; the
			second is a heptapeptide located in the central section of the
			domain.
			Description of pattern(s) and/or profile(s)
			become of patternial analog profile(s)
			Consensus pattern [KR]-P-[PTQ]-[FYLVQH]-S-[FY]-x(2)-[LIVM]-
			x(3,4)-[AC]- [LIM]
			Sequences known to belong to this class detected by the pattern
			ALL, except for AFX1 and FKHR.
			Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern W-[QKR]-[NS]-S-[LIV]-R-H
			Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Last update November 1997 / Patterns and text revised.
			References
			[1]
			Weigel D., Jaeckle H.
			Cell 63:455-456(1990).
			[2]
			Clark K.L., Halay E.D., Lai E., Burley S.K.
			Nature 364:412-420(1993).
			f
			[3]
			Haecker U., Kaufmann E., Hartmann C., Juergens G., Knoechel
			W., Jaeckle H.
			EMBO J. 14:5306-5317(1995).
FtsJ		FtsJ cell division protein	Accession number: PE01709
		· 130 cen division protein	Accession number: PF01728 Definition: FtsJ cell division protein
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B 1791 (release 4.1)
			Gathering cutoffs: -38 -38
			Trusted cutoffs: -20.90 -20.90
			Noise cutoffs: -56.70 -56.70
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 93186701
			Reference Title: The Escherichia coli FtsH protein is a
			prokaryotic member Reference Title: of a protein family of putative ATPases
			Reference Title: of a protein family of putative ATPases involved in
			Reference Title: membrane functions, cell cycle control, and
			gene
			Reference Title: expression.
			Reference Author: Tomoyasu T, Yuki T, Morimura S, Mori H,
			Yamanaka K, Niki H,
			Reference Author: Hiraga S, Ogura T;
			Reference Location: J Bacteriol 1993;175:1344-1351.
			Database Reference INTERPRO; IPR002877;
			Database reference: PFAMB; PB030182;
			Comment: This family consists of FtsJ from various
			bacterial and archaeal sources

		-	896
Pfam	Prosite	Full Name	Description
			Comment: In E. coli FtsJ is not essential for growth but affects cell division [1]. Number of members: 25
FTSW_RODA_SPOVE	PDOC00352	Cell cycle proteins ftsW / rodA / spoVE signature	A number of prokaryotic proteins involved in cell cycle processes have been found [1,2] to be structurally related, these proteins are: - Escherichia coli and related bacteria cell division protein ftsW. This protein plays a role in the stabilization of the ftsZ ring during cell division. - Escherichia coli and related bacteria rod shape-determining protein rodA (or mrdB). It is required for the expression of the enzymatic activity of PBP2, which is thought to participate in the synthesis of peptidoglycan during the initiation of cell elongation. - Bacillus subtilis stage V sporulation protein E (spoVE). The exact function of spoVE in endospore formation is not known. - Bacillus subtilis hypothetical protein ylaO. - Bacillus subtilis hypothetical protein ywcF (ipa-42D). - Cyanophora paradoxa cyanelle ftsW homolog. This protein may be involved in the organelle division process. All these proteins are hydrophobic integral membrane protein and contain about 400 residues. We have selected the best conserved region, which is located in the C-terminal section, as a signature pattern for these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [NV]-x(5)-[GTR]-[LIVMA]-x-P-[PTLIVM]-x-G-[LIVM]-x(3)- [LIVMFW](2)-S-[YSA]-G-G-[STN]-[SA] Sequences known to belong to this class detected by the pattern ALL.
Furin-like			Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Pattern and text revised. References [1] Ikeda M., Sato T., Wachi M., Jung H.K., Ishino F., Kobayashi Y., Matsuhashi M. J. Bacteriol. 171:6375-6378(1989). [2] Joris B., Dive G., Henriques A., Piggot P.J., Ghuysen JM. Mol. Microbiol. 4:513-517(1990).
G. II PIING			Members of this family include receptors that mediate transmembrane signalling. These receptors can bind to a number of factors including: amphiregulin, epidermal growth factor, gp30, heparin-binding egf, insulin, insulin-like growth factor I and II, neuregulins, transforming growth factor-alpha and, and vaccinia virus growth Signal transduction is mediated by catalytic activity of tyrosine kinase, such as ATP + A protein tyrosine = ADP + protein tyrosine phosphate. Typically, such signal transduction have been implicated in metabolic and developmental changes, including cell fate and differentiation. Examples include instruction of follicle cells to follow a dorsal pathway of development rather than the default ventral pathway. may also bind the spitz protein. References describing these family members and their biological activities:

		٤	397
Pfam	Prosite	Full Name	Description Abbot et al., J. Biol. Chem. 267:10759-10763(1992);Araki et al., J. Biol. Chem. 262:16186-16191(1987); Aroian et al., EMBO J. 13:360-366(1994); Aroian et al., Nature 348:693-699(1990); Barbetti et al., Diabetes 41:408-415(1992); Bargmann et al., Nature 319:226-230(1986); Cama et al., J. Biol. Chem. 268:8060-8069(1993); Cama et al., J. Clin. Endocrinol. Metab. 73:894-901(1991); Carrera et al., Hum. Mol. Genet. 2:1437-1441 (1993); Clifford et al., Genetics 137:531-550(1994); Cocozza et al., Diabetes 41:521-526(1992); Cooke et al., Biochem. Biophys. Res. Commun. 177:1113-1120(1991); Coussens et al., Science 230:1132-1139(1985); Dickens et al., Biochem. Biophys. Res. Commun. 186:244-250(1992); Ebina et al., Cell 40:747-758(1985); Ebina et al., Proc. Natl. Acad. Sci. U. S.A. 84:704-708(1987); Ehsani et al., Genomics 15:426-429(1993); Elbein et al., Diabetes 42:429-434(1993); Elbein, Diabetes 38:737-743(1989); Fujita-Yamaguchi et al., Protein Seq. Data Anal. 1:3-6(1987); Gullick et al., EMBO J. 11:43-48(1992); Haruta et al., Diabetes 42:1837-1844(1993); Hubbard et al., EMBO J. 16:5572-5581(1997). Hubbard et al., Nature 372:746-754(1994); Iwanishi et al., Diabetologia 36:414-422(1993); Kadowaki et al., J. Clin. Invest. 86:254-264(1990); Kadowaki et al., Science 240:787-790(1988); Kim et al., Diabetologia 35:261-266(1992); Klinkhamer et al., EMBO J. 8:2503-2507(1989); Kusari et al., J. Biol. Chem. 268:11272-11277(1993); Lee et al., Oncogene 8:3403-3410(1993); Lesokhin et al., Dev. Biol. 205:129-144(1999); Livneh et al., Cell 40:599-607(1985). Longo et al., Proc. Natl. Acad. Sci. U.S.A. 90:60-64(1993); McKeon et al., Mol. Endocrinol. 4:647-656(1990); Moller et al., J. Biol. Chem. 265:14979-14985(1990); Moller et al., Mol. Endocrinol. 4:647-656(1990); Moller et al., J. Biol. Chem. 266:1628-1628-1628-66(1989); Seino et al., Biochem. Biophys. Res. Commun. 189:650-653(1992); Schejter et al., Glochem. 265:1497-14985(1999); Moller et al., J. Biol. Chem. 264:16238-16245(1989); Seino et al., Biochem. Biophys. Res. C
Galactosyl_T		Galactosyltransferase	Accession number: PF01762 Definition: Galactosyltransferase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_885 (release 4.2) Gathering cutoffs: -46 -46 Trusted cutoffs: -43.90 -43.90 Noise cutoffs: -49.80 -49.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 98079080 Reference Title: Cloning of a human Reference Title: UDP-galactose:2-acetamido-2-deoxy-D-glucose 3beta- Reference Title: galactosyltransferase catalyzing the formation of type 1 Reference Author: Reference Author: Reference Number: [2] Reference Number: Reference Medline: 98079027 Reference Title: Genomic cloning and expression of three



		8	398
Pfam	Prosite	Full Name	Description
			murine
			Reference Title: UDP-galactose: beta-N- acetylglucosamine
			Reference Title: beta 1,3-galactosyltransferase genes. Reference Author: Hennet T, Dinter A, Kuhnert P, Mattu TS,
			, , , , , , , , , , , , , , , , , , , ,
			Rudd PM, Berger Reference Author: EG:
			Reference Author: EG; Reference Location: J Biol Chem 1998;273:58-65.
			Database Reference INTERPRO; IPR002659;
			Database reference: PFAMB; PB005938;
			Database reference: PFAMB; PB012965;
			Comment: This family includes the
			galactosyltransferases
		}	Comment: UDP-galactose:2-acetamido-2-deoxy-D-
	ļ		glucose3beta-galactosyltransferase
			Comment: Swiss:O43825 [1] and UDP-Gal:beta-
			GlcNAc beta 1,3-galactosyltranferase
			Comment: Swiss: 054904 [2].
			Comment: Specific galactosyltransferases transfer
			galactose to GlcNAc terminal
			Comment: chains in the synthesis of the lacto-series
			oligosaccharides types 1
			Comment: and 2 [1].
			Number of members: 29
G-alpha		G-protein alpha subunit	Accession number: PF00503
İ			Definition: G-protein alpha subunit
			Author: Finn RD
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_11 (release 1.0)
			Gathering cutoffs: 13.8 13.8
			Trusted cutoffs: 13.80 13.80
			Noise cutoffs: 9.70 12.70
		1	HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1] Reference Medline: 94353239
	1		- 40
1			
			alpha 1 and the Reference Title: mechanism of GTP hydrolysis.
			ME, Gilman AG, Reference Author: Sprang SR;
]	Reference Author: Sprang SR; Reference Location: Science 1994;265:1405-1412.
			Reference Number: [2]
			Reference Medline: 97004345
		i	Reference Title: How G proteins work: a continuing story.
		1	Reference Author: Coleman DE, Sprang SR;
			Reference Location: Trends Biochem Sci 1996;21:41-44.
			Database Reference: PRINTS; PR00318;
			Database Reference: SCOP; 1gia; fa; [SCOP-USA][CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR001019;
			Database Reference PDB; 1gia; 34; 343;
			Database Reference PDB; 1gil; 34; 343;
			Database Reference PDB; 1as0; 32; 344;
		1	Database Reference PDB; 1gfi; 33; 345;
			Database Reference PDB; 1as2; 32; 346;
		1	Database Reference PDB; 1bh2; 32; 346;
			Database Reference PDB; 1cip A; 32; 347;
			Database Reference PDB; 1git; 32; 348;
		İ	Database Reference PDB; 1agr D; 11; 353;
			Database Reference PDB; 1gg2 A; 6; 348;
			Database Reference PDB; 1gp2 A; 6; 348;
		1	Database Reference PDB; 1bof; 10; 353;
			Database Reference PDB; 1as3; 9; 353;
İ		1	Database Reference PDB; 1gdd; 9; 353;
		1	Database Reference PDB; 1agr A; 6; 353;
			Database Reference PDB; 1tag; 27; 340;
			Database Reference PDB; 1tad A; 27; 342;
			Database Reference PDB; 1tad B; 27; 342;
			Database Reference PDB; 1tnd B; 27; 342;
			Database Reference PDB; 1tnd C; 27; 342;
1			Database Reference PDB; 1tad C; 27; 344;
			Database Reference PDB; 1tnd A; 27; 349;

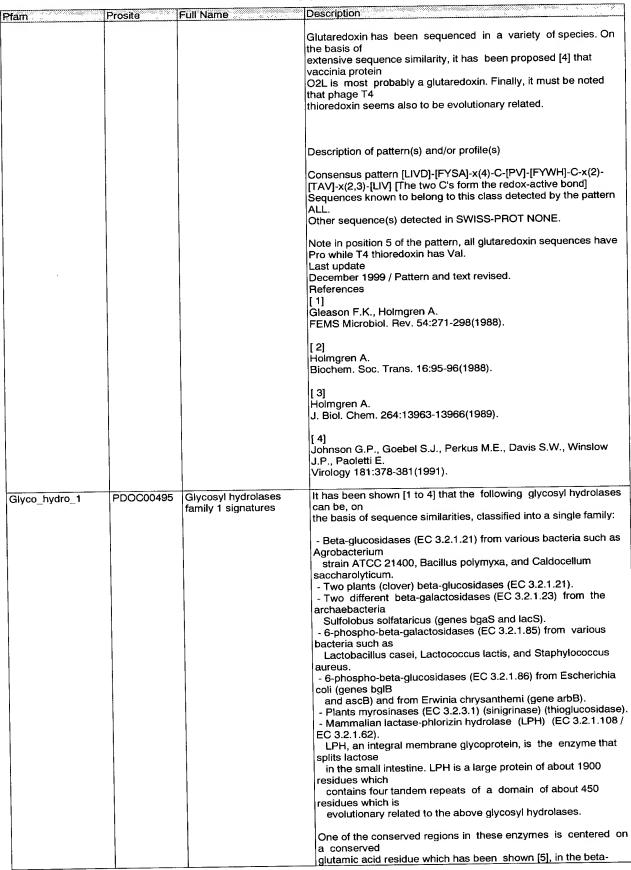
ū	
叮	
Ü	
₽	
Q	
Ū	
Ē	
ä	
O	
H	
W	

4	
4	

Ham	Prosite	Full Name	Description
Pfam Pfam	Prosite	Full Ivaline	Database Reference PDB; 1cjk C; 39; 388;
			Database Reference PDB; 1cjt C; 39; 388;
			Database Reference PDB; 1cju C; 39; 388;
			Database Reference PDB; 1cjv C; 39; 388;
			Database Reference PDB; 1azt A; 35; 391;
			Database Reference PDB; 1azt B; 35; 391;
			Database Reference PDB; 1azs C; 36; 393;
			Database reference: PFAMB; PB034080;
			Comment: G proteins couple receptors of extracellular
			signals to intracellular
	,		Comment: signaling pathways.
			Comment: The G protein alpha subunit binds guanyl
			nucleotide and is a weak
			Comment: GTPase.
			Number of members: 245
GCV H		Glycine cleavage H-	Accession number: PF01597
20 ·		protein	Definition: Glycine cleavage H-protein
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_988 (release 4.1)
			Gathering cutoffs: 25 25
			Trusted cutoffs: 27.90 27.90
			Noise cutoffs: -58.80 -58.80
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 94255425
		1	Reference Title: X-ray structure determination at 2.6-A
			resolution of a
		1	Reference Title: lipoate- containing protein: the H-protein of
			the glycine
			Reference Title: decarboxylase complex from pea leaves.
			Reference Author: Pares S, Cohen-Addad C, Sieker L,
			Neuburger M, Douce R;
			Reference Location: Proc Natl Acad Sci U S A 1994;91:4850-
			4853.
			Database Reference: SCOP; 1htp; fa; [SCOP-USA][CATH-
	1		PDBSUM]
			Database Reference INTERPRO; IPR002930;
			Database Reference PDB; 1hpc A; 2; 127;
			Database Reference PDB; 1hpc B; 2; 127; PDB; 1htp; 2; 127;
			proteins, part of the glycine Comment: cleavage multienzyme complex (GCV)
			Comment: cleavage multienzyme complex (GCV) found in bacteria and the mitochondria
			catabolism of glycine in eukaryotes. Comment: A lipoyl group is attached to a completely
			conserved lysine residue. Comment: The H protein shuttles the methylamine
		1	group of glycine from the
			Comment: P protein to the T protein. Number of members: 40
			Number of members.
GCV T		Glycine cleavage T-	Accession number: PF01571
_	1	protein (aminomethyl	Definition: Glycine cleavage T-protein (aminomethyl
	1	transferase)	transferase)
	1		Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_933 (release 4.0)
		1	Gathering cutoffs: -146 -146
			Trusted cutoffs: -124.50 -124.50
			Noise cutoffs: -167.90 -167.90
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 97199363
			Reference Title: Cloning, and molecular characterization of
			the GCV1 gene
1			Reference Title: encoding the glycine cleavage T-protein
			from Saccharomyces
		1	Reference Title: cerevisiae.

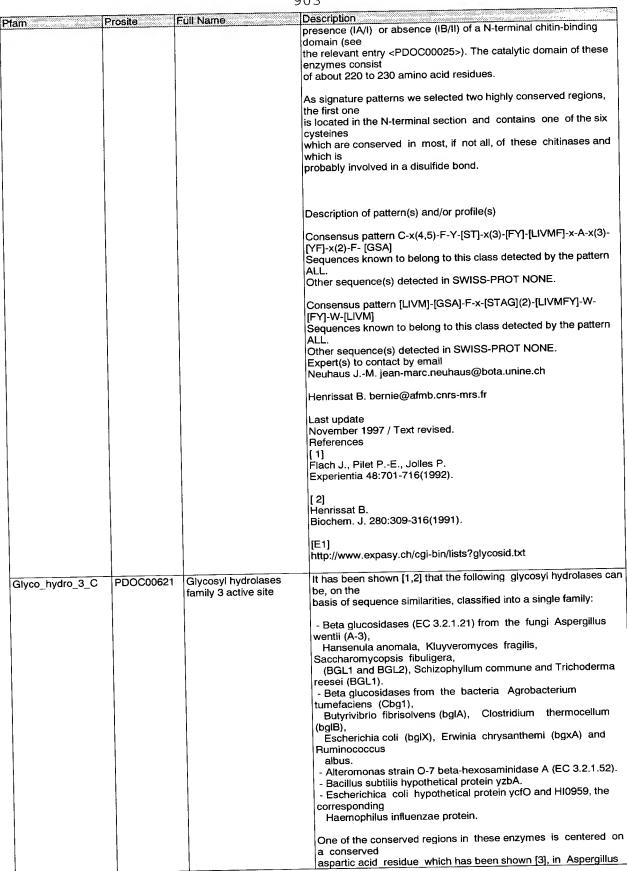


		9	00
Pfam	Prosite	Full Name	Description
			Reference Author: McNeil JB, Zhang F, Taylor BV, Sinclair DA, Pearlman RE, Reference Author: Gene 1997;186:13-20. Database Reference INTERPRO; IPR002536; Database reference: PFAMB; PB004229; Comment: This is a family of glycine cleavage T-proteins, part of the glycine cleavage multienzyme complex (GCV) found in bacteria and the mitochondria Comment: of eukaryotes. GCV catalyses the catabolism of glycine in eukaryotes. Comment: The T-protein is an aminomethyl transferase. Number of members: 27
G-gamma	PDOC01002	G-protein gamma subunit profile	Guanine nucleotide-binding proteins (G proteins) [1] act as intermediaries in the transduction of signals generated by transmembrane receptors. G proteins consist of three subunits (alpha, beta, and gamma). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. The gamma subunits are small proteins (from 70 to 110 residues) that are bound to the membrane via a isoprenyl group (either a farnesyl or a geranyl-geranyl) covalently linked to their C-terminus. In mammals there are at least 12 different isoforms of gamma subunits. The Caenorhabditis elegans protein egl-10, which is a regulator of G-protein signalling, contains a G-protein gamma-like domain. We have developed a profile that spans the complete length of the gamma subunit. Description of pattern(s) and/or profile(s) Sequences known to belong to this class detected by the profile ALL, except for yeast and squid G-protein gamma. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Pennington S.R. srpenn@liverpool.ac.uk Last update November 1997 / First entry. References [1] Pennington S.R. Protein Prof. 2:16-315(1995).
glutaredoxin	PDOC00173	Glutaredoxin	Glutaredoxin [1,2,3], also known as thioltransferase, is a small protein of approximately one hundred amino-acid residues. It functions as an electron carrier in the glutathione-dependent synthesis of deoxyribonucleotides by the enzyme ribonucleotide reductase. Like thioredoxin, which functions in a similar way, glutaredoxin possesses an active center disulfide bond. It exists in either a reduced or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond.





	P	1	302 1=
Pfam	Prosite	Full Name	Description
			glucosidase from Agrobacterium, to be directly involved in glycosidic bond
			cleavage by acting
			as a nucleophile. We have used this region as a signature pattern.
			As a second signature pattern we selected a conserved region, found in the
			N-terminal extremity of these enzymes, this region also contains a glutamic
			acid residue.
			Description of pattern(s) and/or profile(s)
			CO
			Consensus pattern [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G- [LIVMFAR]-[CSAGN] [E is the active site residue] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT 12.
			Note this pattern will pick up the last two domains of LPH; the first two domains, which are removed from the LPH precursor by proteolytic processing, have lost the active site glutamate and may therefore be inactive [4].
			Consensus pattern F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)- [FYNH]-[NQ]-x-E-x- [GSTA] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Note this pattern will pick up the last three domains of LPH. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr
			Last update November 1995 / Patterns and text revised. References
			[1] Henrissat B. Biochem. J. 280:309-316(1991).
			[2] Henrissat B. Protein Seq. Data Anal. 4:61-62(1991).
			[3] Gonzalez-Candelas L., Ramon D., Polaina J. Gene 95:31-38(1990).
			[4] El Hassouni M., Henrissat B., Chippaux M., Barras F. J. Bacteriol. 174:765-777(1992).
			[5] Withers S.G., Warren R.A.J., Street I.P., Rupitz K., Kempton J.B., Aebersold R. J. Am. Chem. Soc. 112:5887-5889(1990).
Glyco_hydro_19	PDOC00620	Chitinases family 19 signatures	Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the
			beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence similarity chitinases belong to either family 18
			or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 19
			(also known as classes IA or I and IB or II) are enzymes from plants that
			function in the defense against fungal and insect pathogens by destroying their chitin-containing cell wall. Class IA/I and IB/II enzymes differ
			in the



			904 Description
efam	Prosite [-ŭll Name	wentii beta- glucosidase A3, to be implicated in the catalytic mechanism. We have used this region as a signature pattern.
			Description of pattern(s) and/or profile(s) Consensus pattern [LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]- [LIVT]-[LIVMF]- [ST]-D-x(2)-[SGADNI] [D is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr Last update November 1997 / Pattern and text revised. References [1] Henrissat B.
			Biochem. J. 280:309-316(1991). [2] Castle L.A., Smith K.D., Morris R.O. J. Bacteriol. 174:1478-1486(1992). [3] Bause E., Legler G. Biochim. Biophys. Acta 626:459-465(1980).
Glyco_hydro_45	PDOC00877	Glycosyl hydrolases family 45 active site	The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family K or as the glycosyl hydrolases family 45 [3,E1]. The enzymes which are currently known to belong to this family are listed below.
			- Endoglucanase 5 from Humicola insolens Endoglucanase 5 from Trichoderma reesei (egl5) Endoglucanase K from Fusarium oxysporum Endoglucanase B from Pseudomonas fluorescens (celB) Endoglucanase 1 from Ustilago maydis (egl1). The best conserved regions in these enzymes is located in the N-terminal section. It contains an aspartic acid residue which has been shown [4] to act as a nucleophile in the catalytic mechanism. We use this region as a signature pattern.
			Description of pattern(s) and/or profile(s) Consensus pattern [STA]-T-R-Y-[FYW]-D-x(5)-[CA] [The D is ar active site residue] Sequences known to belong to this class detected by the patter ALL. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr





Pfam Prosite Full Name Description Last update November 1997 / Pattern and text revised. References [1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990). [2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warr R.A.J. Microbiol. Rev. 55:303-315(1991). [3] Henrissat B., Bairoch A. Biochem. J. 293:781-788(1993). [4] Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Daute Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schum M. Nature 365:362-364(1993). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt	
November 1997 / Pattern and text revised. References [1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990). [2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warr R.A.J. Microbiol. Rev. 55:303-315(1991). [3] Henrissat B., Bairoch A. Biochem. J. 293:781-788(1993). [4] Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Daute Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schulm. N. Nature 365:362-364(1993). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt	
Beguin P. Annu. Rev. Microbiol. 44:219-248(1990). [2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warr R.A.J. Microbiol. Rev. 55:303-315(1991). [3] Henrissat B., Bairoch A. Biochem. J. 293:781-788(1993). [4] Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Daute Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schud M. Nature 365:362-364(1993). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt	
Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warr R.A.J. Microbiol. Rev. 55:303-315(1991). [3] Henrissat B., Bairoch A. Biochem. J. 293:781-788(1993). [4] Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Daute Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schue M. Nature 365:362-364(1993). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt	
Henrissat B., Bairoch A. Biochem. J. 293:781-788(1993). [4] Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Daute Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schur M. Nature 365:362-364(1993). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt	
Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Daute Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schud M. Nature 365:362-364(1993). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt	-
[E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt	
http://www.expasy.ch/cgi-bin/lists?glycosid.txt	
Glyco_hydro_47 Glycosyl hydrolase family Members of this family are alpha-mannosidases that catalys hydrolysis of the terminal 1,2-linked alpha-D-mannose resident the oligo-mannose oligosaccharide Man(9)(GlcNAc)(2). The enzymes are capable of taking part in the glycosylation path and glycoprotein processing.	es in
GTP_cyclohydrol PDOC00672 GTP cyclohydrolase I GTP cyclohydrolase I (EC 3.5.4.16) catalyzes the biosynth formic acid	sis of
and dihydroneopterin triphosphate from GTP. This reaction first step in	s the
the biosynthesis of tetrahydrofolate in prokaryotes, of tetrahydrobiopterin in vertebrates, and of pteridine-containing pigments in insects.	
GTP cyclohydrolase I is a protein of from 190 to 250 amino residues. The	acid
comparison of the sequence of the enzyme from bacterial are ukaryotic sources shows that the structure of this enzyme has been extrem well conserved throughout evolution [1].	İ
As signature patterns we selected two conserved regions. I first contains a perfectly conserved tetrapeptide which is part of the GTP-bi	ŀ
pocket [2], the second region also contains conserved residues involve GTP-binding.	l in
Description of pattern(s) and/or profile(s)	
Consensus pattern [DEN]-[LIVM](2)-x(2)-[KRNQ]-[DEN]-[LIVX(3)-[ST]-x-C-E- H-H Sequences known to belong to this class detected by the pa	,
ALL. Other sequence(s) detected in SWISS-PROT NONE.	
Consensus pattern [SA]-x-[RK]-x-Q-[LIVM]-Q-E-[RN]-[LI]-[TS Sequences known to belong to this class detected by the pa	
Other sequence(s) detected in SWISS-PROT NONE. Last update	
July 1999 / Patterns and text revised. References	
[1] Maier J., Witter K., Guetlich M., Ziegler I., Werner T., Ninner H.	

		3	906
Pfam	Prosite	Full Name	Description
			Biochem. Biophys. Res. Commun. 212:705-711(1995).
			[2] Nar H., Huber R., Meining W., Schmid C., Weinkauf S., Bacher A. Structure 3:459-466(1995).
HCV_capsid		Hepatitis C virus capsid protein	Family members include nucleocapsid proteins of the HCV. This virus family comprises a nnucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein c and mRNA. Uses for these polypeptides include: immunulogical epitopes for vaccines; or as mRNA chaperone proteins to aid in processing or to prevent degradation. References describing examples of these capsid polypeptides include: Chen et al., Virology 188:102-113(1992); and Okamoto et al., J. Gen. Virol. 72:2697-2704(1991)
HD		HD domain	Accession number: PF01966
			Definition: HD domain Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -0.50 Gathering cutoffs: -0.50 -0.50 Noise cutoffs: -0.50 -2.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 99085258 The HD domain defines a new superfamily of metal-dependent Reference Title: phosphohydrolases. Reference Author: Reference Location: Database Reference: Database reference: PFAMB; PB0066725; PFAMB; PB006725; PFAMB; PB009617; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619; PFAMB; PB009619;
HDV_ag		Hepatitis delta virus delta antigen	Accession number: PF01517 Definition: Hepatitis delta virus delta antigen Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_808 (release 4.0) Gathering cutoffs: -8 -8 Trusted cutoffs: 23.30 23.30 Noise cutoffs: -40.50 -40.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 94065676 Characterization of RNA-binding domains of hepatitis delta Reference Title: antigen. Reference Author: Dubois F, Bonelli F, Reference Author: Calogero RA, Goudeau A; Reference Number: [2] Reference Medline: 98362586 Structural basis of the oligomerization of hepatitis delta Reference Title: antigen. Reference Title: Sasses of the oligomerization of hepatitis delta Reference Title: Author: Zuccola HJ, Rozzelle JE, Lemon SM, Erickson BW, Hogle JM; Reference Location: Structure 1998;6:821-830. Database Reference: SCOP; 1a92; fa; [SCOP-USA][CATH-PDBSUM]



Dia	In		90 /
Pfam	Prosite	Full Name	Description INTERPORT INTERPORT
			Database Reference INTERPRO; IPR002506; Database Reference PDB: 1a92 A: 12: 23:
			Database Reference PDB; 1a92 A; 12; 23; Database Reference PDB; 1a92 B; 12; 23;
			Database Reference PDB; 1a92 C; 12; 23;
			Database Reference PDB; 1a92 D; 12; 60;
		1	Database Reference PDB; 1a92 A; 47; 60;
			Database Reference PDB; 1a92 B; 47; 60;
	İ		Database Reference PDB; 1a92 C; 47; 60;
			Comment: The hepatitis delta virus (HDV) encodes a
	1		single protein, the
			Comment: hepatitis delta antigen (HDAg). The central
			region of this protein Comment: has been shown to bind RNA [1] Several
	i		Comment: has been shown to bind RNA [1]. Several interactions are also
			Comment: mediated by a coiled-coil region at the N
			terminus of the protein [2].
			Number of members: 145
hemolysinCabind	PDOC00293	Homolygia type coloium	Commonwhile hard in the state of the state o
nemolysincabina	PD0C00293	Hemolysin-type calcium- binding region signature	Gram-negative bacteria produce a number of proteins which are secreted into
		amang region eignature	the growth medium by a mechanism that does not require a
			cleaved N-terminal
			signal sequence. These proteins, while having different functions,
			seem [1] to
			share two properties: they bind calcium and they contain a
			variable number of
			tandem repeats consisting of a nine amino acid motif rich in glycine, aspartic
			acid and asparagine. It has been shown [2] that such a domain
			is involved in
			the binding of calcium ions in a parallel beta roll structure. The
			proteins
			which are currently known to belong to this category are:
			- Hemolysins from various species of bacteria. Bacterial
			hemolysins are
			exotoxins that attack blood cell membranes and cause cell
			rupture. The
			hemolysins which are known to contain such a domain are
			those from: E. coli
			(gene hlyA), A. pleuropneumoniae (gene appA), A.
			actinomycetemcomitans
			and P. haemolytica (leukotoxin) (gene lktA).
			- Cyclolysin from Bordetella pertussis (gene cyaA). A multifunctional protein
			which is both an adenylate cyclase and a hemolysin.
			- Extracellular zinc proteases: serralysin (EC 3.4.24.40) from
]		Serratia, prtB
			and prtC from Erwinia chrysanthemi and aprA from
			Pseudomonas aeruginosa.
			- Nodulation protein nodO from Rhizobium leguminosarum.
			We derived a signature pattern from conserved positions in the
			sequence of the
			calcium-binding domain.
			
			Description of pattern(s) and/or profile(s)
			= ====================================
	İ		Consensus pattern D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D
			Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Note this pattern is found once in nodO and the extracellular
			proteases but up to 5 times in some hemolysin/cyclolysins.
			Last update
			October 1993 / Text revised.
			References
			[1]
			Economou A., Hamilton W.D.O., Johnston A.W.B., Downie J.A.
			EMPO 1 0:040 0E4/4000)
			EMBO J. 9:349-354(1990).

Pfam Pr	osite Full Name				
		[2]			
		Baumann U., Wu S., Flaherty K.M., McKay D.B. EMBO J. 12:3357-3364(1993).			
Heptosyltranf	Heptosyltransferase	Accession number: PF01075			
	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	Definition: Heptosyltransferase			
		Author: Finn RD, Bateman A			
		Alignment method of seed: Clustalw			
		Source of seed members: Pfam-B_839 (release 3.0)			
		Gathering cutoffs: -40 -40			
		Trusted cutoffs: -31.80 -31.80			
		Noise cutoffs: -47.10 -47.10			
		HMM build command line: hmmbuild -F HMM SEED			
		HMM build command line: hmmcalibrateseed 0 HMM			
		Reference Number: [1]			
		Reference Medline: 98112827			
		Reference Title: Enzymatic synthesis of lipopolysaccharide in			
		Escherichia			
1		Reference Title: coli. Purification and properties of			
		heptosyltransferase I. Reference Author: Kadrmas JL, Raetz CR;			
İ					
		Reference Location: J Biol Chem 1998;273:2799-2807.			
		Database Reference INTERPRO; IPR002201;			
		Database reference: PFAMB; PB021100;			
		Database reference: PFAMB; PB033445;			
		Database reference: PFAMB; PB041423;			
		Comment: Lipopolysaccharide is a major component of			
		the outer leaflet of			
		Comment: the outer membrane in Gram-negative			
i		bacteria. It is composed of			
		Comment: three domains; lipid A, Core oligosaccharide and the O-antigen.			
		Comment: All of these enzymes transfer heptose to the lipopolysaccharide			
		Comment: core.			
		Number of members: 46			
Herpes alk exo	Herpesvirus alkaline	Accession number: PF01771			
	exonuclease	Definition: Herpesvirus alkaline exonuclease			
İ		Author: Bashton M, Bateman A			
		Alignment method of seed: Clustalw			
		Source of seed members: Pfam-B_822 (release 4.2)			
		Gathering cutoffs: 25 25			
•		Trusted cutoffs: 318.00 318.00			
		Noise cutoffs: -277.60 -277.60			
		HMM build command line: hmmbuild -F HMM SEED			
		HMM build command line: hmmcalibrateseed 0 HMM			
		Reference Number: [1]			
	1				
1		Reference Medline: 85107093			
1		Reference Medline: 85107093 Reference Title: Studies on the herpes simplex virus alkaline			
		Reference Title: Studies on the herpes simplex virus alkaline			
		Reference Title: Studies on the herpes simplex virus alkaline nuclease:			
		Reference Title: Studies on the herpes simplex virus alkaline nuclease: Reference Title: detection of type-common and type-specific			
		Reference Title: Studies on the herpes simplex virus alkaline nuclease: Reference Title: detection of type-common and type-specific epitopes on the			
		Reference Title: Studies on the herpes simplex virus alkaline nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme.			
		Reference Title: nuclease: Reference Title: epitopes on the Reference Title: Reference Author: Studies on the herpes simplex virus alkaline detection of type-common and type-specific enzyme. Banks LM, Halliburton IW, Purifoy DJ,			
		Reference Title: Studies on the herpes simplex virus alkaline nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell			
		Reference Title: Studies on the herpes simplex virus alkaline nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Banks LM, Halliburton IW, Purifoy DJ, KL;			
		Reference Title: Studies on the herpes simplex virus alkaline nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Banks LM, Halliburton IW, Purifoy DJ, Killington RA, Powell Reference Author: KL; Gen Virol 1985;66:1-14.			
		Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Author: Reference Location: Database Reference Studies on the herpes simplex virus alkaline detection of type-common and type-specific enzyme. Banks LM, Halliburton IW, Purifoy DJ, KL; J Gen Virol 1985;66:1-14. INTERPRO; IPR001616;			
		Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Author: Reference Location: Database Reference Comment: Studies on the herpes simplex virus alkaline on the herpes simplex virus alkaline of type-common and type-specific enzyme. Banks LM, Halliburton IW, Purifoy DJ, KL; J Gen Virol 1985;66:1-14. INTERPRO; IPR001616; This family includes various alkaline			
		Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Author: Reference Location: Database Reference Comment: exonucleases from Studies on the herpes simplex virus alkaline enzyme. detection of type-common and type-specific enzyme. Banks LM, Halliburton IW, Purifoy DJ, KL; J Gen Virol 1985;66:1-14. INTERPRO; IPR001616; This family includes various alkaline			
		Reference Title: nuclease: Reference Title: epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Location: Database Reference Comment: cyonucleases from Comment: muclease: Studies on the herpes simplex virus alkaline detection of type-common and type-specific enzyme. Banks LM, Halliburton IW, Purifoy DJ, KL; J Gen Virol 1985;66:1-14. INTERPRO; IPR001616; This family includes various alkaline members of the herpesviridae. Alkaline			
		Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Location: Database Reference Comment: exonucleases from Comment: exonuclease			
		Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Location: Database Reference Comment: exonucleases from Comment: exonuclease Comment: appears to have an important role in the replication of Comment: herpes simplex virus [1].			
		Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Location: Database Reference Comment: exonucleases from Comment: exonuclease Comment: appears to have an important role in the resonance and type-specific detection of type-common and type-specific enzyme. Banks LM, Halliburton IW, Purifoy DJ, KL; J Gen Virol 1985;66:1-14. INTERPRO; IPR001616; This family includes various alkaline members of the herpesviridae. Alkaline appears to have an important role in the			
Herpes_gl	Alphaherpesvirus	Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Location: Database Reference Comment: exonucleases from Comment: exonuclease Comment: exonuclease Comment: exonuclease Comment: herpes simplex virus [1]. Number of members: Studies on the herpes simplex virus alkaline enzyme. Banks LM, Halliburton IW, Purifoy DJ, KL; J Gen Virol 1985;66:1-14. INTERPRO; IPR001616; This family includes various alkaline exonuclease appears to have an important role in the replication of Comment: herpes simplex virus [1].			
Herpes_gl	Alphaherpesvirus glycoprotein I	Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Location: Database Reference Comment: exonucleases from Comment: exonuclease Comment: enzyme. Banks LM, Halliburton IW, Purifoy DJ, KL; J Gen Virol 1985;66:1-14. INTERPRO; IPR001616; This family includes various alkaline exonuclease Comment: appears to have an important role in the replication of Comment: herpes simplex virus [1].			
Herpes_gl	•	Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Location: Database Reference Comment: exonucleases from Comment: exonuclease Comment: exonuclease Comment: exonuclease Comment: herpes simplex virus [1]. Number of members: PF01688			
Herpes_gl	•	Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Author: Reference Author: Killington RA, Powell Reference Author: Reference Location: Database Reference Comment: exonucleases from Comment: exonuclease Comment: exonuclease Comment: exonuclease Comment: herpes simplex virus [1]. Number of members: PF01688 Definition: Alphaherpesvirus glycoprotein I Author: Bashton M, Bateman A Alignment method of seed: Clustalw			
Herpes_gl	•	Reference Title: nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Killington RA, Powell Reference Author: Reference Location: Database Reference Comment: exonucleases from Comment: exonuclease Comment: exonuclease Comment: exonuclease Comment: appears to have an important role in the replication of Comment: 23 Accession number: PF01688 Definition: Alphaherpesvirus glycoprotein I Author: Bashton M, Bateman A			

		9	09
Pfam	Prosite	Full Name	Description
			Trusted cutoffs: 157.20 157.20
			Noise cutoffs: -126.70 -126.70
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			Reference Number: [1] Reference Medline: 96357074
			Reference Title: Biosynthesis of glycoproteins E and I of
			feline
			Reference Title: herpesvirus: gE-gl interaction is required for
			Reference Title: intracellular transport.
			Reference Author: Mijnes JD, van der Horst LM, van Anken
			E, Horzinek MC, Reference Author: Rottier PJ, de Groot RJ;
			Reference Author: Rottier PJ, de Groot RJ; Reference Location: J Virol 1996;70:5466-5475.
		1	Reference Number: [2]
			Reference Medline: 94267406
			Reference Title: Identification of the feline herpesvirus type 1
			(FHV-1)
			Reference Title: genes encoding glycoproteins G, D, I and E:
			expression of Reference Title: FHV-1 glycoprotein D in vaccinia and
		1	raccoon poxviruses.
	1		Reference Author: Spatz SJ, Rota PA, Maes RK;
			Reference Location: J Gen Virol 1994;75:1235-1244.
			Reference Number: [3]
			Reference Medline: 94267879
			Reference Title: Unusual phosphorylation sequence in the
			gpIV (gl) component Reference Title: of the varicella-zoster virus gpl-gpIV
			glycoprotein complex
			Reference Title: (VZV gE-gl complex).
			Reference Author: Yao Z, Grose C;
			Reference Location: J Virol 1994;68:4204-4211.
			Database Reference INTERPRO; IPR002874;
			Comment: This family consists of glycoprotein I form
		İ	various members of the Comment: alphaherpesvirinae these include
		İ	Comment: alphanerpesvirinae these include herpesvirus, varicella-zoster virus
			Comment: and pseudorables virus. Glycoprotein I (gl)
			is important during natural
			Comment: infection, mutants lacking gl produce smaller
			lesions at the site of
	İ		Comment: infection and show reduced neuronal spread
			[1]. gl forms a heterodimeric Comment: complex with gE; this complex displays Fc
			receptor activity (binds to
			Comment: the Fc region of immunoglobulin) [1].
			Glycoproteins are also important
			Comment: in the production of virus-neutralizing
			antibodies and cell mediated
			Comment: immunity [2]. The alphaherpesvirinae have a
			dsDNA gnome and have no Comment: RNA stage during viral replication.
			Number of members: 22
Herpes glycop_D		Herpesvirus glycoprotein	Accession number: PF01528
3.7.6.2.2.3.7.7.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2	1	М	Definition: Herpesvirus glycoprotein M
			Author: Bateman A
			Alignment method of seed: Clustalw Source of seed members: Pfam-B_929 (release 4.0)
			Gathering cutoffs: 25 25
			Trusted cutoffs: 197.30 197.30
	1		Noise cutoffs: -229.70 -229.70
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 96357105 Reference Title: Identification and characterization of
			Reference Title: Identification and characterization of pseudorables virus
			Reference Title: glycoprotein gM as a nonessential virion
			component.
			Reference Author: Dijkstra JM, Visser N, Mettenleiter TC,
			Klupp BG;
			Reference Location: J Virol 1996;70:5684-5688.

		9	10
Pfam	Prosite	Full Name	Description [6]
			Reference Number: [2] Reference Medline: 95381611 Reference Title: of the murine Reference Title: cytomegalovirus homolog of the human cytomegalovirus UL100 Reference Title: gene. Reference Author: Li W, Eidman K, Gehrz RC, Kari B; Reference Location: Virus Res 1995;36:163-175. Database Reference Comment: The herpesvirus glycoprotein M (gM) is an integral membrane protein Comment: predicted to contain 8 transmembrane segments [2]. Glycoprotein M is Comment: not essential for viral replication [1]. Number of members: 24
HesB-like	PDOC00887	Hypothetical hesB/yadR/yfhF family signature	The following uncharacterized proteins have been shown [1] to share regions of similarities:
			 Anabaena and related cyanobacteria protein hesB which may be required for nitrogen fixation. Escherichia coli hypothetical protein yadR and HI1723, the corresponding Haemophilus influenzae protein. Escherichia coli hypothetical protein ydiC. Escherichia coli hypothetical protein yfhF and HI0376, the corresponding Haemophilus influenzae protein. Mycobacterium tuberculosis hypothetical protein Rv2204c. Synechocystis strain PCC 6803 hypothetical protein slr1417. Synechocystis strain PCC 6803 hypothetical protein slr1565. A hypothetical protein in the nifU 5'region of many nitrogen fixing bacteria. Porphyra purpurea chloroplast hypothetical protein in apcF-rps4 intergenic region. Yeast hypothetical protein YLL027W. Yeast hypothetical protein YPR067W. These are small proteins (106 to 135 amino-acid residues in bacteria, about 200 residues in fungi) that contain a number of conserved regions. The most noteworthy of these regions is located in the C-terminal extremity, it contains two conserved cysteines. We have used this region as a signature pattern.
			Description of pattern(s) and/or profile(s) Consensus pattern F-x-[LIVMFY]-x-N-[PG]-[NSKQ]-x(4)-C-x-C-[GS]-x-S-F
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Bairoch A., Rudd K.E. Unpublished observations (1995).
HisG	PDOC01020	ATP phosphoribosyltransferas e signature	ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we



		9	911
Pfam	Prosite	Full Name	Description
			selected a region located in the C-terminal part of this enzyme.
			Description of pattern(s) and/or profile(s)
			Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]- G-x-T-[LM]
			Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / First entry.
histone	PDOC00045 PDOC00046	Histone H2A signature; Histone H4 signature;	Histone H2A is one of the four histones, along with H2B, H3 and H4, which
	PDOC00040 PDOC00287 PDOC00308	Histone H3 signatures; Histone H2B signature	forms the eukaryotic nucleosome core. Using alignments of histone H2A
		· ···orionio i i i i i originariari	sequences [1,2,E1] we selected, as a signature pattern, a conserved region in
			the N-terminal part of H2A. This region is conserved both in classical S-
			phase regulated H2A's and in variant histone H2A's which are synthesized throughout the cell cycle.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [AC]-G-L-x-F-P-V Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT 2.
			Last update November 1995 / Pattern and text revised. References
			[1] Wells D.E., Brown D.
			Nucleic Acids Res. 19:2173-2188(1991).
			[2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).
			[E1] http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES/index.html
			Histone H4 is one of the four histones, along with H2A, H2B and H3, which forms the eukaryotic nucleosome core. Along with H3, it plays a
			central role in nucleosome formation. The sequence of histone H4 has
			remained almost invariant in more then 2 billion years of evolution [1,E1]. The
			region we use as a signature pattern is a pentapeptide found in positions 14 to
			18 of all H4 sequences. It contains a lysine residue which is often acetylated
			[2] and a histidine residue which is implicated in DNA-binding [3].
			Description of pattern(s) and/or profile(s)
			Consensus pattern G-A-K-R-H Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 1.
			Last update November 1995 / Text revised.

	\	S	912
Pfam	Prosite	Full Name	Description
·			References [1] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).
			[2] Doenecke D., Gallwitz D. Mol. Cell. Biochem. 44:113-128(1982).
			[3] Ebralidse K.K., Grachev S.A., Mirzabekov A.D. Nature 331:365-367(1988).
			[E1] http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES/index.html
			Histone H3 is one of the four histones, along with H2A, H2B and H4, which forms the eukaryotic nucleosome core. It is a highly conserved protein of 135 amino acid residues [1,2,E1].
			The following proteins have been found to contain a C-terminal H3-like domain:
			- Mammalian centromeric protein CENP-A [3]. Could act as a core histone necessary for the assembly of centromeres Yeast chromatin-associated protein CSE4 [4] Caenorhabditis elegans chromosome III encodes two highly related proteins (F54C8.2 and F58A4.3) whose C-terminal section is evolutionary related to the last 100 residues of H3. The function of these proteins is not yet known.
			We developed two signature patterns, The first one corresponds to a perfectly conserved heptapeptide in the N-terminal part of H3. The second one is derived from a conserved region in the central section of H3.
			Description of pattern(s) and/or profile(s)
			Consensus pattern K-A-P-R-K-Q-L Sequences known to belong to this class detected by the pattern ALL, except for the H3-like proteins and some protozoan H3. Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern P-F-x-[RA]-L-[VA]-[KRQ]-[DEG]-[IV] Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References
			[1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991).
			[2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).
			[3] Sullivan K.F., Hechenberger M., Masri K. J. Cell Biol. 127:581-592(1994).
			[4] Stoler S., Keith K.C., Curnick K.E., Fitzgerald-Hayes M.



			913
Pfam	Prosite	Full Name	Description
			Genes Dev. 9:573-586(1995).
			[E1]
			http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES/index.html
			Uistans LIOD is one of the four historic clans with LIOA LIO
			Histone H2B is one of the four histones, along with H2A, H3
			and H4, which forms the eukaryotic nucleosome core. Using alignments of
			histone H2B
			sequences [1,2,E1], we selected a conserved region in the C-
			terminal part of
			H2B.
		:	
			Description of pattern(s) and/or profile(s)
			Consensus pattern [KP] E [LIVM] [EO] T v/2\ [KP]-v-[LIVM](2\-v-
			Consensus pattern [KR]-E-[LIVM]-[EQ]-T-x(2)-[KR]-x-[LIVM](2)-x- [PAG]-[DE]-L- x-[KR]-H-A-[LIVM]-[STA]-E-G
			Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Last update
			November 1995 / Pattern and text revised.
			References
			[1]
			Wells D.E., Brown D.
			Nucleic Acids Res. 19:2173-2188(1991).
			r 02
			[2]
			Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).
			Nucleic Acids Res. 22.174-179(1994).
			[E1]
			http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES/index.html
			3 - ,
HMA	PDOC00804		
	FD000004	Heavy-metal-associated	A conserved domain of about 30 amino acid residues has been
	FDCC00804	domain	A conserved domain of about 30 amino acid residues has been found [1] in a
	FDCC00804	I	found [1] in a number of proteins that transport or detoxify heavy metals.
	FD000004	I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain
	FB0000804	I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the
	FBCC00804		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of
	FD000004		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-
	FD000004		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has
	FD000004		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-
	FD000004		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in:
	FB000004		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has
	FB0000804		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are</pdoc00139>
	FB0000004		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in</pdoc00139>
	F D C C C C C C C C C C C C C C C C C C		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both</pdoc00139>
	F D C C C C C C C C C C C C C C C C C C		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies</pdoc00139>
	- DOC00804		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding</pdoc00139>
	F D C C C C C C C C C C C C C C C C C C		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from</pdoc00139>
	F D C C C C C C C C C C C C C C C C C C		found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain</pdoc00139>
			found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the</pdoc00139>
			found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid p1258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixl</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixl from Rhizobium</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixl from Rhizobium meliloti, pacS from Synechococcus strain PCC 7942),</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixl from Rhizobium meliloti, pacS from Synechococcus strain PCC 7942), Mycobacterium leprae</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixl from Rhizobium meliloti, pacS from Synechococcus strain PCC 7942), Mycobacterium leprae ctpA and ctpB and Escherichia coli hypothetical protein yhhO.</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixI from Rhizobium meliloti, pacS from Synechococcus strain PCC 7942), Mycobacterium leprae ctpA and ctpB and Escherichia coli hypothetical protein yhhO. In all these</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixl from Rhizobium meliloti, pacS from Synechococcus strain PCC 7942), Mycobacterium leprae ctpA and ctpB and Escherichia coli hypothetical protein yhhO.</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixl from Rhizobium mellioti, pacS from Synechococcus strain PCC 7942), Mycobacterium leprae ctpA and ctpB and Escherichia coli hypothetical protein yhhO. In all these ATPases the HMA domain(s) are located in the N-terminal</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixl from Rhizobium meliloti, pacS from Synechococcus strain PCC 7942), Mycobacterium leprae ctpA and ctpB and Escherichia coli hypothetical protein yhhO. In all these ATPases the HMA domain(s) are located in the N-terminal section.</pdoc00139>
		I	found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <pdoc00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixI from Rhizobium meliloti, pacS from Synechococcus strain PCC 7942), Mycobacterium leprae ctpA and ctpB and Escherichia coli hypothetical protein yhhO. In all these ATPases the HMA domain(s) are located in the N-terminal section. - Mercuric reductase (EC 1.16.1.1) (gene merA) which is</pdoc00139>

·	1=		914
Pfam	Prosite	Full Name	Description
			bacteria. Mercuric reductase is a class-1 pyridine nucleotide-disulphide oxidoreductase (see
			<pdoc00073>). There is generally one HMA domain (with the exception of a</pdoc00073>
			chromosomal merA from Bacillus strain RC607 which has
			two) in the N- terminal part of merA.
			- Mercuric transport protein periplasmic component (gene merP), also encoded
			by plasmids carried by mercury-resistant Gram-negative bacteria. It seems
			to be a mercury scavenger that specifically binds to one Hg(2+) ion and
			which passes it to the mercuric reductase via the merT protein. The N-
			terminal half of merP is a HMA domain Helicobacter pylori copper-binding protein copP.
			- Yeast protein ATX1 [2], which could act in the transport and/or
			partitioning of copper.
			The consensus pattern for HMA spans the complete domain.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIVNS]-x(2)-[LIVMFA]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]-x(3)-[LIV]-x(9,11)-[IVA]-x-[LVFYS] [The two C's
			probably bind metals] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT 6.
			Last update December 1999 / Pattern and text revised.
			References
			Bull P.C., Cox D.W. Trends Genet. 10:246-252(1994).
			[2]
			Lin SJ., Culotta V.L.
			Proc. Natl. Acad. Sci. U.S.A. 92:3784-3788(1995).
HMG-CoA_red	PDOC00064	Hydroxymethylglutaryl- coenzyme A reductase	Hydroxymethylglutaryl-coenzyme A reductase (EC 1.1.1.34) (HMG-CoA reductase)
		signatures and profile	[1,2] catalyzes the NADP-dependent synthesis of mevalonate from 3-hydroxy-3-
			methylglutaryl-CoA. In vertebrates, HMG-CoA reductase is the rate-limiting
			enzyme in cholesterol biosynthesis. In plants, mevalonate is the precursor of
			all isoprenoid compounds.
			HMG-CoA reductase is a membrane bound enzyme. Structurally, it consists of 3
			domains. An N-terminal region that contains a variable number of transmembrane
			segments (7 in mammals, insects and fungi; 2 in plants), a linker region and a
			C-terminal catalytic domain of approximately 400 amino-acid residues.
			In archebacteria [3] HMG-CoA reductase, which is involved in the
			biosynthesis of the isoprenoids side chains of lipids, seems to be cytoplasmic
			and lack the N-terminal hydrophobic domain.
			Some bacteria, such as Pseudomonas mevalonii, can use mevalonate as the sole
			carbon source. These bacteria use an NAD-dependent
			HMG-CoA reductase



(EC 1.1.1.8) to deacetylate mevalonate into 3-hydroxy-3-methytigulanty-CoA. [3]. The Pseudomonas enzyme is structurally related to the catalytic domain of NADP-dependent HMG-GoA reductases. We selected: three conserved regions as signature patterns for HMG-GoA reductases. The first is located in the center of the catalytic domain, the second is a glycine-rich region located in the C-terminal section of the same catalytic domain and the third is also located in the C-terminal section and catalytic domain and the third is also located in the C-terminal section and catalytic domain and the third is also located in the C-terminal section and catalytic domain and the third is also located in the C-terminal section and catalytic domain and the third is also located in the C-terminal section and catalytic domain and the third is also located in the C-terminal section and catalytic domain and the third is also located in the C-terminal section and catalytic domain and the third is also located in the C-terminal section and catalytic domain and the third is also located in the C-terminal section and catalytic domain and catalytic domain and the third catalytic domain and ca	Pfam	Prosite	Full Name	Description
methylgiutany-CoA [3]. The Pseudomonas enzyme is structurally related to the catasytic domain of NADP-dependent HMG-CoA reductases. We selected three conserved regions as signature patterns for HMG-CoA reductases. The first is located in the custer of the catalytic domain, the second is a glycine-rich region located in the C-terminal section catalytic domain and the third is also located in the C-terminal section and contains an institution and section and contains an institution region located in the C-terminal section and contains an institution region of pattern (s) and/or profile(s) Description of pattern(s) and/or profile(s) Consensus pattern (RKH]-V(S)-DX.M.GX.NX.LVIMA) Sequences known to belong to this class detected by the pattern ALL LLL Consensus pattern (RIVM]-GX.LIVM]-GA.CA.TX.LVIMA] Sequences known to belong to this class detected by the pattern of the sequence(s) detected in SWISS-PROT 4. Consensus pattern A.L.LIVM]-GA.LA.TX.DLIJX-KIRNO-I/GSA]-H.LIMX-FYTU-II] is an active set residual; by the pattern of the sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. November 1997 / Patterns and text revised; profile added. References 1 11 Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Bol. 13,627-638(1989). 1 2] Bascon M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 43:797-3008(1989). 1 3] Lam W.L., Dooltitle W.F. J. Biol. Chem. 267:5829-5834(1992). 1 4] Bacch M.J., Rodwell V.W. J. Bacteriol. 171:2594-3001 (1989). 5 5 Juny R.M. A. May P. Rodwell V.W. J. Biol. Chem. 267:5924-15070(1982). 5 Juny R.M. A. Wang Y., Rodwell V.W. J. Biol. Chem. 267:5924-15070(1982). 5 Juny R.M. A. Wang Y., Rodwell V.W. J. Cenzyme A lysas active site. A pattern and a profile and active pattern and a profile and active pattern and a profile		1.10010	1301130	
catalytic domain of NADP-dependent HMG-CoA reductases. We selected: three conserved regions as signature patterns for hMG-CoA reductases. The first is located in the center of the catalytic domain, the second is a glycine-rich region located in the C-terminal section of the same catalytic domain and the third is also located in the C-terminal section and institute residue that seems [4] to be implicated in the catalytic domain and the third is also located in the C-terminal section and institute residue that seems [4] to be implicated in the catalytic mechanism as a general base. Description of pattern(s) and/or profile(s) Consensus pattern [RKH]-x(s)-D-x-M-G-x-N-x-(LVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern A-(LVM)-x-(LVM)-G-G-IAG)-T Sequences known to belong to this class detected by the pattern ALL sequences known to belong to this class detected by the pattern ALL sequences known to belong to this class detected by the pattern ALL sequences known to belong to this class detected by the pattern ALL sequences known to belong to this class detected by the pattern ALL sequences known to belong to this class detected by the pattern ALL sequences are considered by the pattern ALL sequences are considered by the pattern ALL sequences are considered by the pattern ALL sequences are considered by the profile ALL. Consensus pattern A-(LVM)-x-(STAM)-x(2)-(LV)-x-(KRNO)-(GSA)-H-(LM)-x-(FV-H-)-[H is an active site residue) Sequences known to belong to this class detected by the pattern ALL sequences are considered by the profile ALL. Consensus pattern A-(LVM)-x-(STAM)-x(2)-(LV)-x-(KRNO)-(FV-X-X-X-(LVM)-x-(STAM)-x(2)-(LV-X-X-X-X-(LVM)-x-(STAM)-x(2)-(LV-X-X-X-X-(LVM)-x-(STAM)-x(2)-(LV-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X				methylglutaryl-CoA
of NADP-dependent HMG-CoA reductases. We selected three conserved regions as signature patterns for HMG-CoA textuctases. The first is located in the center of the catalytic second is a glycine-rich region located in the C-terminal section of the same catalytic domain and the third is also located in the C-terminal section and contains an histidine residue that seems [4] to be implicated in the catalytic mechanism as a general base. Description of pattern(s) and/or profile(s) Consensus pattern [RKH]-x(6)-D-x-M-G-x-N-x-(LVMA) Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern and a sequence(s) detected in SWISS-PROT 5. Consensus pattern and a sequence(s) detected in SWISS-PROT MONE. Sequences known to belong to this class detected by the pattern ALL. White is an active site residual sequences and the sequences are sequences as a sequence and the sequence and the sequence and the sequence a				
We selected: three conserved regions as signature patterns for MMG-CoA reductases. The first is located in the center of the catalytic second is a glycine-rich region located in the C-terminal section of the same catalytic domain and the third is also located in the C-terminal section and contains an histoline residue that seems [4] to be implicated in the catalytic mechanism as a general base. Description of pattern(s) and/or profile(s) Description of pattern(s) and/or profile(s) Consensus pattern [RivF1, xi6]-D.x-M.Gx-N.x-[LVMA]. Sequences innown to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LVM]-G-x-[LVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. ALL. Consensus pattern [ALVM]-G-x-[LVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Consensus pattern [VT-I] [FI is an active site residue] Sequences known to belong to this class detected by the profile ALL. Consensus pattern [VT-I] [FI is an active site residue] Sequences known to belong to this class detected by the profile ALL. Consensus pattern [VT-I] [FI is an active site residue] Sequences known to belong to this class detected by the profile ALL. Consensus pattern [VT-I] [FI is an active site residue] Sequences known to belong to this class detected by the profile ALL. Consensus pattern [VT-I] [FI is an active site residue] Sequences known to belong to this class detected by the profile ALL. Consensus pattern [VT-I] [FI is an active site residue] Sequences known to belong to this class detected by the profile ALL. Consensus pattern [VT-I] [FI is an active site residue] S		1		
for HMG-CoA reductases. The first is located in the center of the catalytic domain, the second is a glycine-rich region located in the C-terminal section of the same section and in and the third is also located in the C-terminal section and contains an histidine residue that seems [4] to be implicated in the catalytic mechanism as a general base. Description of pattern(s) and/or profile(s) Consensus pattern [FikH]-x(s)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-x-[LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Consensus pattern A-[LIVM]-G-X-[LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-X-[STAN]-x(g-ILI)-x-[KRNO]-[GSA]- H-[LM]-x-[FV-II]-II is an active set residue) Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the penting and the profile is much more sensitive than the penting and the pattern and profile and profile and profile and profile. As the profile is much more sensitive than the penting and profile a				
reductases. The first is located in the center of the catalytic domain, the second is a glycine-rich region located in the C-terminal section of the same catalytic domain and the third is also located in the C-terminal section and contains an histidine residue that seems [4] to be implicated in the contains an instidine residue that seems [4] to be implicated in the contains an instidine residue that seems [4] to be implicated in the contains as a general base. Description of pattern(e) and/or profitie(s) Consensus pattern [RiKH]-x(6)-D-xM-G-x-N-x(LIVMA] Sequences known to belong to this class detected by the pattern ALL. Consensus pattern [RiKH]-x(6)-D-xM-G-x-N-x(LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-S-Z-LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile in author have sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References Calcies C., Ferrer A., Balcelle L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 6:3797-3808(1989). [3] Lam W.L., Doolitile W.F. J. Biol. Chem. 2677:5064-15070(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 17:12994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 17:12994-3001(1989). [6] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 17:12994-3001(1989).				
domain, the second is a glycine-rich region located in the C-terminal section of the same catalytic domain and the third is also located in the C-terminal section and contains an insidine residue that seems [4] to be implicated in the catalytic domain and the third is also located in the C-terminal section and contains an insidine residue that seems [4] to be implicated in the catalytic domain as a general base. Description of pattern(s) and/or profile(s) Consensus pattern [RKH]-K(6)-D-x-M-G-x-N-K[LVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LVM]-G-3-[J-M]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LVM]-X-STAN]-X(2)-[LI]-X-KRINO]-(GSA)-H-[LM]-X-[PVL1]-[H is an active site residue) Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use if if you have access to the necessary software tools to do so. Union of the sequence(s) detected in SWISS-PROT NONE. References [1] Calles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 6:3797-3806(1988). [3] Lam W.L., Docititle W.F. J. Biol. Chem. 267:15064-15070(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Damay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). 1MGL-like PDOC00813 PDOC00843 PDOC00843 PDOC00843				
of the same catalytic domain and the third is also located in the C-terminal section and contains an histidine residue that seems [4] to be implicated in the catalytic mechanism as a general base. Description of pattern(s) and/or profile(s) Consensus pattern [RKH]-x(s)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LVMA]-G-x-[LVM]-G-x-[LVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LVM]-x-[STAN]-x(2)-[LI]-x-[KRNO]-(GSA)-H-[LM]-x-[FYLH]-[H] is an active site residue] Sequences known to belong to this class detected by the pattern ALL. except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update Norman and profile. As the profile is much more sensitive than the pattern with the pattern with the pattern of the profile added. References I 11 Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:797-380(1989). [3] Lam W.L., Dooilttle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [6] Light-is-propylimalate All. (Co.A. 1, 4) calcetocacteat in Light-is-propylimalate All-proposed and profile. All-proposed and profile. All-proposed and profile. All-proposed and profile. All-proposed and profile. All-proposed and profile. All-proposed and profile. All-proposed and profile. All				domain, the
catalytic domain and the third is also located in the C-terminal section and contains an histidine residue that seems [4] to be implicated in the catalytic mechanism as a general base. Description of pattern(s) and/or profile(s) Consensus pattern [RiKH]+x(6)-D-x-M-G-x-N-x-[LVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-x-[LVM]-G-x-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A_LIVM]-x-(STAN)-x(S-ILM]-x-(RNO)-(GSA)-H-LIMA-x-[PX-ILH]-th is an article with set residual) Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the north sequence of the profile is much more sensitive than the north sequence of the profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensitive than the north profile is much more sensi				
contains an histidine residue that seems [4] to be implicated in the catalytic mechanism as a general base. Description of pattern(s) and/or profile(s) Consensus pattern [RKH-]-x(6)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-x-[LIVM]-G-a-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-G-x-[LIVM]-G-a-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-x-[STAN-x(2)-[LI)-x-[KRNO]-[GSA]-H-[M-M]-x-[P*[LI*]-[H is an active site residue] search ALL. except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary set of so. Last update to do so. Lost update t				
catalytic mechanism as a general base. Description of pattern(s) and/or profile(s) Consensus pattern [RKH]-x(6)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-x-[LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-x-[STAN]-x(2)-[LI)-x-[KRNO]-[GSA]-H-[LM]-x-[FYH-I] [H is an active site residue) Sequences known to belong to this class detected by the pattern ALL, except for aronabeabcrist IMG-COA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13.627-638(1989). [2] Basson M.E., Thorsness M., Finer-Morer J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1989). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Boach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5064-15070(1992). **MGIL-like** PDOC00643 **Hydroxymethylglutaryl-toenzyme A lyase (HMG-CoA lyase or HL) (ECC 4.1.3.4) coenzyme A lyase active site; Alpha-isopropylmalate Alpha-isopropylmalate Alpha-isopropylmalate Alpha-isopropylmalate Alpha-isopropylmalate				
Description of pattern(s) and/or profile(s) Consensus pattern [RKH]-x(6)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-x-[LVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern A-[LIVM]-G-x-[LVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1999). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1998). [3] Lam W.L., Doolitie W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Basach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992).				
Consensus pattern [FKH]-x(6)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-x-[LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-x-[STAM]-x(2)-[LI]-x-[KRNO]-[GSA]-H-[LM]-x-[FYLH] [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use if if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1989). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 171:2994-3001(1999). [6] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992).				
Consensus pattern [FKH]-x(6)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-x-[LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-x-[STAM]-x(2)-[LI]-x-[KRNO]-[GSA]-H-[LM]-x-[FYLH] [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use if if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1989). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 171:2994-3001(1999). [6] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992).				
Consensus pattern [FKH]-x(6)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-x-[LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-x-[STAM]-x(2)-[LI]-x-[KRNO]-[GSA]-H-[LM]-x-[FYLH] [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use if if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1989). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 171:2994-3001(1999). [6] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992).			Ì	
Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[X-IVM]-X-[X-IVM]-X				Description of pattern(s) and/or profile(s)
Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-[LIVM]-G-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[Q-X-IVM]-X-[STAN]-X-[X-IVM]-X-[X-IVM]-X				Concensus nettern [DKH] v/C) D v M C v N v [I I)/MAA
ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-X-[LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-X-[STAN]-X/2]-[LI]-X-[RNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-X-[FXNO]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-X-[FXNO]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-X-[FXNO]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-X-[FXNO]-[III]-X-[FXNO]-[III]-X-[FXNO]-[III]-X-[FXNO]-[GSA]- H-[LM]-X-[FXLH]-X-[FXNO]-[III]-X-[IXNO]-[III]-X-[IXNO]-[III]-X-[IXNO]-[III]-X-[IXNO]-[III]-X-[IXNO]-[III]-X-[II				
Consensus pattern [LIVM]-G-x-[LIVM]-G-C-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-x-[STAN]-x(2)-[LI]-x-[KRNO]-[GSA]-H-[LM]-x-[FYLH] [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13-627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 6:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Boronomyme A lyase active site; Alpha-isopropylmalate Apha-isopropylmalate				ALL.
Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LI/M]-x-[STAN]-x(2)-[LI]-x-[KRNQ]-[GSA]-H-[LM]-x-[FYLH] [It is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13.627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). 3-hydroxy-3-methyligilutaryl-coenzyme A lyase (HMG-CoA lyase or HU) (EC 4-13.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transfor				Other sequence(s) detected in SWISS-PROT 4.
Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LI/M]-x-[STAN]-x(2)-[LI]-x-[KRNQ]-[GSA]-H-[LM]-x-[FYLH] [It is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13.627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). 3-hydroxy-3-methyligilutaryl-coenzyme A lyase (HMG-CoA lyase or HU) (EC 4-13.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transfor				Consensus pattern [LIVM]-G-x-[LIVM]-G-G-[AG]-T
Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-x-[STAN]-x(2)-[LI]-x-[KRNQ]-[GSA]-H-[LM]-x-[FYL-H] [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [11] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13.627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). 4MGL-like PDOC00813 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or Hu) (CC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and acitalyzes the transformation of HMG-CoA into acetyl-CoA and				
Consensus pattern A-[LIVM]-x-[STAN]-x(2)-[LI]-x-[KRNO]-[GSA]-H-[LM]-x-[FYLH] [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5529-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). 3-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and active-cloacetate. In				
H-[LM]-x - [FYL-H] [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) caltayzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				office sequence(s) detected in SWIGG-1 110 1 3.
Sequences known to belong to this class defected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 171:2994-3001(1992). 3-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catelyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetale. in				
ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. Jibid. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				
Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). IMGL-like PDOC00813 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) a-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA and acetoacetate. In				
ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). 4MGL-like PDOC00813 PDOC00643 PDOC00643 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or thus) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				Other sequence(s) detected in SWISS-PROT NONE.
ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). 4MGL-like PDOC00813 PDOC00643 PDOC00643 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or thus) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				Sequences known to belong to this class detected by the profile
Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). [4] Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) sale; site; Alpha-isopropylmalate				ALL.
and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). [4] Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) coenzyme A lyase active site; Alpha-isopropylmalate				Other sequence(s) detected in SWISS-PROT NONE.
pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [11] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). [4MGL-like PDOC00813 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PHydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. in				Note this documentation entry is linked to both a signature pattern
software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 PHydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. in				
Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doollittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 PDOC00643 PDOC00644 PDOC00645 Hydroxymethylglutaryl- coenzyme A lyase active site; Alpha-isopropylmalate Alpha-isopropylmalate				
References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). Hydroxymethylglutaryl-coenzyme A lyase active site; Alpha-isopropylmalate References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1989). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). 3-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				Last update
[1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). [4] Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HU) (EC 4.1.3.4) site; Alpha-isopropylmalate				
Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). Hydroxymethylglutaryl-coenzyme A lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				1
[2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) coenzyme A lyase active site; Alpha-isopropylmalate				Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A.
Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				Plant Mol. Biol. 13:627-638(1989).
Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				[2]
[3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J.
Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 Hydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				Moi. Cell. Biol. 8:3797-3808(1988),
J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 Hydroxymethylglutaryl- coenzyme A lyase active site; Alpha-isopropylmalate Alpha-isopropylmalate J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). 3-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				
[4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 Hydroxymethylglutaryl- coenzyme A lyase active site; Alpha-isopropylmalate All (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				· ·
Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PHydroxymethylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) Catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				J. Dioi. Criem. 267:5829-5834(1992).
J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PHDOC00				
Topic February F				
Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 Hydroxymethylglutaryl- coenzyme A lyase active site; Alpha-isopropylmalate Algorithm and process of the proce				J. Dauteriol. 171.2994-3001(1969).
J. Biol. Chem. 267:15064-15070(1992). HMGL-like PDOC00813 PDOC00643 PDOC00				
HMGL-like PDOC00813 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 PDOC00643 Site; Coenzyme A lyase active site; Alpha-isopropylmalate Alpha-isopropylmalate Sa-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) Catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				
PDOC00643 coenzyme A lyase active site; HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In				J. DIGI. CHEIT. 207.10004-10070(1992).
site; catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In	HMGL-like			
Alpha-isopropylmalate acetoacetate. In		PDOC00643		
		Ì		vertebrates it is a mitochondrial enyme which is involved in

	·		916
Pfam	Prosite	Full Name	Description
		synthases signatures	ketogenesis and in leucine catabolism [1]. In some bacteria, such as
			Pseudomonas mevalonii,
			it is involved in mevalonate catabolism (gene mvaB). A cysteine has been shown
			[2], in mvaB, to be required for the activity of the enzyme. The
			region around this residue is perfectly conserved and is used as a signature
			pattern.
			Description of pattern(a) and (an another to
			Description of pattern(s) and/or profile(s)
			Consensus pattern S-V-A-G-L-G-G-C-P-Y [C is the active site residue]
			Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update
	İ		November 1995 / First entry. References
			[1]
			Mitchell G.A., Robert MF., Hruz P.W., Wang S., Fontaine G., Behnke C.E., Mende-Mueller L.M., Schappert K., Lee C., Gibson
			K.M., Miziorko H.M.
			J. Biol. Chem. 268:4376-4381(1993).
;			[2] Hruz P.W., Narasimhan C., Miziorko H.M.
			Biochemistry 31:6842-6847(1992).
			The following enzymes have been shown [1] to be functionally
			as well as
			evolutionary related:
			- Alpha-isopropylmalate synthase (EC 4.1.3.12) which catalyzes
			the first step in the biosynthesis of leucine, the condensation of acetyl-CoA
			and alpha- ketoisovalerate to form 2-isopropylmalate synthase.
			- Homocitrate synthase (EC 4.1.3.21) (gene nifV) which is
			involved in the biosynthesis of the iron-molybdenum cofactor of nitrogenase
			and catalyzes
			the condensation of acetyl-CoA and alpha-ketoglutarate into homocitrate.
			- Soybean late nodulin 56.
			- Methanococcus jannaschii hypothetical proteins MJ0503,
			MJ1195 and MJ1392.
			We have selected two conserved regions as signature
	ļ		patterns for these enzymes. The first region is located in the N-terminal section
			while the second region is located in the central section and contains two
			conserved
			histidine residues which could be implicated in the catalytic mechanism.
			Description of pattern(s) and/or profile(s)
			Consensus pattern L-R-[DE]-G-x-Q-x(10)-K
			Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern [LIVMFW]-x(2)-H-x-H-[DN]-D-x-G-x-[GAS]-x-
			[GASLI]
			Sequences known to belong to this class detected by the pattern ALL.
	L		



Pfam	Prosite	Full Name	Pescription
r tani	riosite	UIENCLIES	Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Wang SZ., Dean D.R., Chen JS., Johnson J.L.
hormone5	PDOC00237	Neurohypophysial hormones signature	J. Bacteriol. 173:3041-3046(1991). Oxytocin (or ocytocin) and vasopressin [1] are small (nine amino acid residues), structurally and functionally related neurohypophysial peptide hormones. Oxytocin causes contraction of the smooth muscle of the uterus and of the mammary gland while vasopressin has a direct antidiuretic action on the kidney and also causes vasoconstriction of the peripheral vessels. Like the majority of active peptides, both hormones are synthesized as larger protein precursors that are enzymatically converted to their mature forms. Peptides belonging to this family are also found in birds, fish, reptiles and amphibians (mesotocin, isotocin, valitocin, glumitocin, aspargtocin, vasotocin, seritocin, asvatocin, phasvatocin), in worms (annetocin), octopi (cephalotocin), locust (locupressin or neuropeptide F1/F2) and in molluscs (conopressins G and S) [2]. The pattern developed to detect this category of peptides spans their entire sequence and includes four invariant amino acid residues.
			Description of pattern(s) and/or profile(s) Consensus pattern C-[LIFY](2)-x-N-[CS]-P-x-G [The two C's are linked by a disulfide bond]. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Pattern and text revised. References [1] Acher R., Chauvet J. Biochimie 70:1197-1207(1988). [2] Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R. Int. J. Pept. Protein Res. 45:482-487(1995).
НРРК	PDOC00631	7,8-dihydro-6- hydroxymethylpterin- pyrophosphokinase signature	All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Most microorganisms must synthesize folate de novo because they lack the active transport system of higher vertebrate cells which allows these organisms to use dietary folates. Enzymes involved in folate biosynthesis are therefore targets for a variety of antimicrobial agents such as trimethoprim or sulfonamides. 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (EC 2.7.6.3) (HPPK) catalyzes the attachment of pyrophosphate to 6-hydroxymethyl-7,8-dihydropterin to form 6-hydroxymethyl-7,8-dihydropteridine pyrophosphate.



			918
Pfam	Prosite	Full Name	Description
			This is the first step in a three-step pathway leading to 7,8-dihydrofolate.
			Bacterial HPPK (gene folk or sulD) [1] is a protein of 160 to 270 amino
			acids. In the lower eukaryote Pneumocystis carinii, HPPK is the central domain
			of a multifunctional folate synthesis enzyme (gene fas) [2].
			As a signature for HPPK, we selected a conserved region located in the central section of these enzymes.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [KRHD]-x-[GA]-[PSAE]-R-x(2)-D-[LIV]-D- [LIVM](2) Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update
			July 1999 / Pattern and text revised. References [1]
			Talarico T.L., Ray P.H., Dev I.K., Merrill B.M., Dallas W.S. J. Bacteriol. 174:5971-5977(1992).
			[2] Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves C.J.
			Gene 112:213-218(1992).
HTH_AraC	PDOC00040	Bacterial regulatory proteins, araC family signature and profile	The many bacterial transcription regulation proteins which bind DNA through a 'helix-turn-helix' motif can be classified into subfamilies on the
			basis of sequence similarities. One of these subfamilies groups together the following proteins [1,2]:
			- aarP, a transcriptional activator of the 2'-N-acetyltransferase gene in Providencia stuartii.
			- ada, an Escherichia coli and Salmonella typhimurium bifunctional protein that repairs alkylated guanine in DNA by transferring the alkyl
			group at the O(6) position to a cysteine residue in the enzyme. The
			methylated protein acts a positive regulator of its own synthesis and of the alkA,
			alkB and aidB genes adaA, a Bacillus subtilis bifunctional protein that acts both as a
			transcriptional activator of the ada operon and as a methylphosphotriester-
			DNA alkyltransferase adiY, an Escherichia coli protein of unknown function aggR, the transcriptional activator of aggregative adherence fimbria I
			expression in enteroaggregative Escherichia coli appY, a protein which acts as a transcriptional activator of acid
			phosphatase and other proteins during the deceleration phase of growth and acts as a repressor for other proteins that are synthesized in
			exponential growth or in the stationary phase.
			- araC, the arabinose operon regulatory protein, which activates the transcription of the araBAD genes.
			- cafR, the Yersinia pestis F1 operon positive regulatory protein.



Pfam	Prosite	Full Name Description
TIGHTS WAS A SECOND	r rosite	- celD, the Escherichia coli cel operon repressor.
		- cfaD, a protein which is required for the expression of the CFA/I adhesin
		of enterotoxigenic Escherichia coli.
		- csvR, a transcriptional activator of fimbrial genes in
		enterotoxigenic
		Escherichia coli envY, the porin thermoregulatory protein, which is involved in
		the control
		of the temperature-dependent expression of several
		Escherichia coli envelope proteins such as ompF, ompC, and lamB.
		- exsA, an activator of exoenzyme S synthesis in Pseudomonas
		aeruginosa.
		- fapR, the positive activator for the expression of the 987P operon coding
		for the fimbrial protein in enterotoxigenic Escherichia coli.
		- hrpB, a positive regulator of pathogenicity genes in
		Burkholderia
		solanacearum invF, the Salmonella typhimurium invasion operon regulator.
		- marA, which may be a transcriptional activator of genes
		involved in the
		multiple antibiotic resistance (mar) phenotype meIR, the melibiose operon regulatory protein, which
		activates the
		transcription of the melAB genes.
		- mixE, a Shigella flexneri protein necessary for secretion of ipa
		invasins mmsR, the transcriptional activator for the mmsAB operon in
		Pseudomonas
		aeruginosa.
		- msmR, the multiple sugar metabolism operon transcriptional
		activator in Streptococcus mutans.
		- pchR, a Pseudomonas aeruginosa activator for pyochelin and
		ferripyochelin
		receptor perA, a transcriptional activator of the eaeA gene for
		intimin in
		enteropathogenic Escherichia coli.
		- pocR, a Salmonella typhimurium regulator of the cobalamin
• 101		biosynthesis operon.
		- parA, from Proteus vulgaris.
		- rafR, the regulator of the raffinose operon in Pediococcus
		pentosaceus. - ramA, from Klebsiella pneumoniae.
		- rhaR, the Escherichia coli and Salmonella typhimurium L-
		rhamnose operon
		transcriptional activator rhaS, an Escherichia coli and Salmonella typhimurium positive
		activator of
		genes required for rhamnose utilization.
		- rns, a protein which is required for the expression of the cs1
		and cs2 adhesins of enterotoxigenic Escherichia coli.
		- rob, a protein which binds to the right arm of the replication
		origin oriC
		of the Escherichia coli chromosome.
		- soxS, a protein that, with the soxR protein, controls a superoxic
		response regulon in Escherichia coli.
		- tetD, a protein from transposon TN10.
		- tcpN or toxT, the Vibrio cholerae transcriptional activator of
		the tcp operon involved in pilus biosynthesis and transport.
		- thcR, a probable regulator of the the operon for the
		degradation of the
		thiocarbamate herbicide EPTC in Rhodococcus sp. strain
		NI86/21 ureR, the transcriptional activator of the plasmid-encoded urea
1	1	operon in
1		Enterobacteriaceae.



	1		920
Pfam	Prosite	Full Name	Description
			activator.
			- virF, the Shigella transcriptional factor of invasion related antigens
			ipaBCD.
			- xyIR, the Escherichia coli xylose operon regulator xyIS, the transcriptional activator of the Pseudomonas putida
			TOL plasmid
			(pWWO, pWW53 and pDK1) meta operon (xyIDLEGF genes).
			- yfeG, an Escherichia coli hypothetical protein. - yhiW, an Escherichia coli hypothetical protein.
			- yhiX, an Escherichia coli hypothetical protein.
			- yidL, an Escherichia coli hypothetical protein. - yijO, an Escherichia coli hypothetical protein.
			- yuxC, a Bacillus subtilis hypothetical protein.
			- yzbC, a Bacillus subtilis hypothetical protein.
			Except for celD, all of these proteins seem to be positive
			transcriptional factors. Their size range from 107 (soxS) to 529 (yzbC) residues.
			The helix-turn-helix motif is located in the third quarter of most
			of the
			sequences; the N-terminal and central regions of these proteins are presumed
			to interact with effector molecules and may be involved in dimerization [3].
			The minimal DNA binding domain, which spans roughly 100 residues and comprises
			the HTH motif contains another region with similarity to classical
			HTH domain. However, it contains an insertion of one residue in the turn-
			region.
			A signature pattern was derived from the region that follows the
			first HTH domain and that includes the totality of the putative second HTH
			domain. A
			more sensitive detection of members of the araC family is available through
			the use of a profile which spans the minimal DNA-binding
			region of 100 residues.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [KRQ]-[LIVMA]-x(2)-[GSTALIV]-{FYWPGDN}-
			x(2)-[LIVMSA]- x(4,9)-[LIVMF]-x(2)-[LIVMSTA]-[GSTACIL]-x(3)-
			[ĠÁŇQRF]- [ĹĬVMFÝ]-x(4,5)-[ĹFÝ]-x(3)-[FYIVÁ]-{FYWHCM}-x(3)- [GSADENQKR]-x-[NSTAPKL]-[PARL]
			Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT 37.
			Sequences known to belong to this class detected by the profile
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the
			pattern, you should use it if you have access to the necessary
			software tools to do so. Expert(s) to contact by email
			Ramos J.L. jlramos@samba.cnb.uam.es
			Gallegos MT. mtrini@samba.cnb.uam.es
			Last update
			November 1997 / Text revised. References
			Gallegos MT., Michan C., Ramos J.L.

			921
Pfam	Prosite	Full Name	Description
			Nucleic Acids Res. 21:807-810(1993).
		İ	[2]
			Henikoff S., Wallace J.C., Brown J.P.
			Meth. Enzymol. 183:111-132(1990).
		1	
			[3]
		i	Bustos S.A., Schleif R.F.
			Proc. Natl. Acad. Sci. U.S.A. 90:5638-5642(1993).
Hydrolase		haloacid dehalogenase-	Accession number: PF00702
-		like hydrolase	Definition: haloacid dehalogenase-like hydrolase
			Author: Bateman A
			Alignment method of seed: Clustalw
		i	Source of seed members: Pfam-B_566 (release 2.1)
			Trusted cutoffs: 7.10 7.10
			Noise cutoffs: 2.90 2.90
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
İ			Reference Number: [1]
			Reference Medline: 96355356
			Reference Title: Crystal structure of L-2-haloacid
			dehalogenase from
			Reference Title: Pseudomonas sp. YL. An alpha/beta
			hydrolase structure that
			Reference Title: is different from the alpha/beta hydrolase
			fold.
			Reference Author: Hisano T, Hata Y, Fujii T, Liu JQ, Kurihara
i			T, Esaki N,
			Reference Author: Soda K;
			Reference Location: J Biol Chem 1996;271:20322-20330.
			Database Reference: SCOP; 1jud; sf; [SCOP-USA][CATH-
1			PDBSUM]
i			Database Reference INTERPRO; IPR001454;
			Database Reference PDB; 1jud; 4; 197;
			Database Reference PDB; 1zrm; 4; 197;
			Database Reference PDB; 1zrn; 4; 197;
1			Database Reference PDB; 1aq6 A; 2; 193;
			Database Reference PDB; 1aq6 B; 2; 193;
			Database Reference PDB; 1qq5 A; 2; 193;
			la
			LEE/ 1997 T/T/ 1997
			, ,
			Database Reference PDB; 1qq7 B; 2; 193;
			Database Reference PDB; 1cqz A; 4; 19;
			Database Reference PDB; 1cr6 A; 4; 19;
			Database Reference PDB; 1cqz B; 4; 206;
			Database Reference PDB; 1cr6 B; 4; 206;
	ļ		Database Reference PDB; 1cqz A; 48; 206;
	}		Database Reference PDB; 1cr6 A; 48; 206;
			Database reference: PFAMB; PB000701;
]		Database reference: PFAMB; PB001048;
	Ì		Database reference: PFAMB; PB019234;
			Database reference: PFAMB; PB032787;
	1		Database reference: PFAMB; PB040985;
	1		Database reference: PFAMB; PB041061;
		,	Database reference: PFAMB; PB041182;
]		Database reference: PFAMB; PB041477;
	1		Database reference: PFAMB; PB041535;
	I		Database reference: PFAMB; PB041628;
	-		Database reference: PFAMB; PB041677;
			Comment: This family are structurally different from the
	l		alpha/
	İ		Comment: beta hydrolase family (abhydrolase).
			Comment: Deta hydrolase family (abhydrolase). Comment: This family includes L-2-haloacid
			dehalogenase, epoxide
			Comment: hydrolases and phosphatases.
			Comment: The structure of the family consists of two
			domains. One



Yam	Prosite	Full Name	Description	and a sign of the dignment hetween
			Comment:	conserved region of the alignment, between
			residues 16 and	96 of Swiss:P24069. The rest of the fold is
			Comment:	96 of Swiss:P24069. The rest of the fold is
			composed of the	core alpha/beta domain.
			Comment: Number of members	
			Iddiliper of members	. 104
lypB_UreG		HypB/UreG nucleotide-	Accession number:	PF01495
ypb_orca		binding domain		lypB/UreG nucleotide-binding domain
]		Author: B	ashton M, Bateman A
			Alignment method of	seed: Clustalw
				bers: Pfam-B_428 (release 4.0)
			Gathering cutoffs:	25 25
			Trusted cutoffs:	197.70 197.70
			Noise cutoffs:	-40.00 -40.00 d line: hmmbuild -F HMM SEED
			HIMM build comman	d line: hmmcalibrateseed 0 HMM
			Reference Number:	[1]
			Reference Medline:	97285753
			Reference Title:	The HypB protein from Bradyrhizobium
			japonicum can store	
			Reference Title:	nickel and is required for the nickel-
			dependent	
			Reference Title:	transcriptional regulation of hydrogenase.
			Reference Author:	Olson JW, Fu C, Maier RJ;
			Reference Location:	Mol Microbiol 1997;24:119-128.
		1	Reference Number:	[2]
			Reference Medline:	97352660
		1	Reference Title:	Characterization of UreG, identification of a
	1	ì	Reference Title:	UreD-UreF-UreG complex, and
			evidencesuggesting	nucleotide-binding site in UreG is required
			Reference Title:	nucleotide-binding site in ored is required
			for in vivo	metallocenter assembly of Klebsiella
			Reference Title:	metallocenter assembly of responding
			aerogenes urease. Reference Author:	Moncrief MB, Hausinger RP;
			Reference Location	
	1		Reference Number:	
			Reference Medline:	
			Reference Title:	The product of the hypB gene, which is
			required for nickel	ino product of the cylindrical
			Reference Title:	incorporation into hydrogenases, is a nove
			guanine	
			Reference Title:	nucleotide-binding protein.
			Reference Author:	Maier T, Jacobi A, Sauter M, Bock A;
			Reference Location	
		1	Reference Number	
			Reference Medline:	92325016
			Reference Title:	Klebsiella aerogenes urease gene cluster
			sequence of ureD	Lateral and the tour appearant
			Reference Title:	and demonstration that four accessory
			genes (ureD, ureE,	ureF, and ureG) are involved in nickel
			Reference Title:	urer, and ured) are involved in theker
	1	İ	metallocenter	biosynthesis.
			Reference Title: Reference Author:	Lee MH, Mulrooney SB, Renner MJ,
			Markowicz Y, Haus	
			Reference Location	
			Database Reference	
		1	Comment:	This domain is found in HypB, a
			hydrogenase expre	
		i	Comment:	protein, and UreG a urease accessory
			protein. Both these	e proteins contain
			Comment:	a P-loop nucleotide binding motif [2,3].
		1	HypB has GTPase	activity
	1		Comment:	and is a guanine nucleotide binding protei
			[3]. It is not known	_
			Comment:	whether UreG binds GTP or some other
	1		nucleotide. Both e	nzvmes are involved
			Comment:	in nickel binding. HypB can store nickel ar
	1		is required for nick	el
			Comment:	dependent hydrogenase expression [1].
			UreG is required for	or functional
1	1	1	Comment:	incorporation of the urease nickel

itani.	_	7
	i	-
din.		7
thur.	Į	¥
	Ë	200
	Ī	7.4
Tim!		-
Hall	_	1
2		
	=	1
init	==	-
•	4	
	=	
-	=	=
1	_	1

923				
Pfam	Prosite	Full Name	Description	
			metallocenter.[4] GTP hydrolysis may Comment: required by these proteins for nickel	
			incorporation into other nickel	
			Comment: proteins [1].	
			Number of members: 41	
			PE01740	
IBB		Importin beta binding	Accession number: PF01749 Definition: Importin beta binding domain	
		domain	Definition: Importin beta binding domain Author: Bashton M, Bateman A	
			Alignment method of seed: Clustalw	
			Source of seed members: Pfam-B_544 (release 4.2)	
			Gathering cutoffs: 25 25	
			Trusted cutoffs: 67.30 67.30	
			Noise cutoffs: -15.90 -15.90 HMM build command line: hmmbuild -F HMM SEED	
			HMM build command line: hmmcalibrateseed 0 HMM	
			Reference Number: [1]	
			Reference Medline: 98359119	
			Reference Title: Crystallographic analysis of the recognition	
			of a nuclear	
			Reference Title: localization signal by the nuclear import	
			factor Reference Title: karyopherin alpha.	
1		1	Reference Author: Conti E, Uy M, Leighton L, Blobel G,	
			Kuriyan J;	
		1	Reference Location: Cell 1998;94:193-204.	
			Reference Number: [2]	
			Reference Medline: 98275030 Reference Title: Importins and exportins: how to get in and	
			out of the	
			Reference Title: nucleus [published erratum appears in	
			Trends Biochem Sci	
			Reference Title: 1998 Jul;23(7):235]	
			Reference Author: Weis K;	
			Reference Location: Trends Biochem Sci 1998;23:185-189. Reference Number: [3]	
	•		Reference Medline: 98250643	
			Reference Title: Transport into and out of the cell nucleus.	
			Reference Author: Gorlich D;	
			Reference Location: EMBO J 1998;17:2721-2727.	
			Reference Number: [4]	
			Reference Medline: 96270582 Reference Title: The binding site of karyopherin alpha for	
			karyopherin beta	
			Reference Title: overlaps with a nuclear localization	
			sequence.	
			Reference Author: Moroianu J, Blobel G, Radu A;	
			Reference Location: Proc Natl Acad Sci U S A 1996;93:6572-	
			6576. Reference Number: [5]	
			Reference Medline: 96203101	
			Reference Title: A 41 amino acid motif in importin-alpha	
			confers binding to	
			Reference Title: importin- beta and hence transit into the	
			nucleus. Reference Author: Gorlich D, Henklein P, Laskey RA,	
			Hartmann E;	
			Reference Location: EMBO J 1996;15:1810-1817.	
			Database Reference: SCOP; 1bk5; fa; [SCOP-USA][CATH-	
			PDBSUM]	
			Database Reference INTERPRO; IPR002652;	
			Database Reference PDB; 1ejl I; 72; 99; Database Reference PDB; 1ejy I; 72; 99;	
			Database Reference PDB; 1ial A; 44; 99;	
			Database Reference PDB; 1qgr B; 28; 51;	
			Database Reference PDB; 1qgk B; 11; 54;	
1			Database Reference PDB; 1ee5 A; 90; 110;	
			Database Reference PDB; 1bk5 A; 89; 110;	
			Database Reference PDB; 1bk5 B; 89; 110; PDB; 1bk6 A; 89; 110;	
			Database Reference PDB; 1bk6 A; 89; 110; Database Reference PDB; 1bk6 B; 89; 110;	
			Database Reference PDB; 1ee4 A; 87; 110;	
			Database Reference PDB; 1ee4 B; 87; 110;	
			Comment: This family consists of the importin alpha	



		· · · · · · · · · · · · · · · · · · ·	924
Pfam	Prosite	Full Name	Description
			(karyopherin alpha),
	1		Comment: importin beta (karyopherin beta) binding domain. The domain mediates
			Comment: formation of the importin alpha beta
			complex; required for classical
		1	Comment: NLS import of proteins into the nucleus,
			through the nuclear pore
		1	Comment: complex and across the nuclear envelope.
			Comment: Also in the alignment is the NLS of importin alpha which overlaps
			Comment: with the IBB domain [4].
			Number of members: 38
IF-2B		Initiation factor 2 subunit	Accession number: PF01008
		family	Definition: Initiation factor 2 subunit family
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_1302 (release 3.0)
			Gathering cutoffs: -135 -135 Trusted cutoffs: -82.40 -82.40
			Noise cutoffs: -02.40 -02.40
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 98188271
			Reference Title: Archaeal translation initiation revisited: the
	<u> </u>		initiation Reference Title: factor 2 and eukarvotic initiation factor 2B
			Reference Title: factor 2 and eukaryotic initiation factor 2B alpha-beta-delta subunit families.
			Reference Author: Kyrpides NC, Woese CR;
			Reference Location: Proc Natl Acad Sci U S A 1998;95:3726-
			3730.
			Database Reference INTERPRO; IPR000649;
			Comment: This family includes initiation factor 2B
			alpha, beta and delta
			Comment: subunits from eukaryotes, initiation factor 2B
			subunits 1 and 2 Comment: from archaebacteria and some proteins of
			Comment: from archaebacteria and some proteins of unknown function from
			Comment: prokaryotes. Initiation factor 2 binds to Met-
			tRNA, GTP and the
			Comment: small ribosomal subunit.
			Number of members: 33
IF3	PDOC00723	Initiation factor 3	Initiation factor 3 (IF-3) (gene infC) [1] is one of the three
		signature	factors required for the initiation of protein biosynthesis in bacteria. IF-
			3 is
			thought to function as a fidelity factor during the assembly of the
			initiation complex which consist of the 30S ribosomal subunit,
			the initiator
			tRNA and the messenger RNA. IF-3 binds to the 30S ribosomal
			subunit; it is a basic protein of 141 to 212 residues.
			The chloroplast initiation factor IE 2/obil in a protein that
			The chloroplast initiation factor IF-3(chl) is a protein that enhances the
	}		poly(A,U,G)-dependent binding of the initiator tRNA to
			chloroplast ribosomal 30s subunits. In its mature form it is a protein of about 400
			residues whose
			central section is evolutionary related to the sequence of bacterial IF-3 [2].
			As a signature pattern we selected a highly conserved region located in the
			central section of bacterial IF-3 and of IF-3(chl).
			Description of pattern(s) and/or profile(s)
			Consensus pattern [KR]-[LIVM](2)-[DN]-[FY]-[GSN]-[KR]-



		3	925
Pfam	Prosite	Full Name	Description [LIVMFYS]-x-[FY]- [DEQTH]-x(2)-[KRQ] Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References
			[1] Liveris D., Schwartz J.J., Geertman R., Schwartz I. FEMS Microbiol. Lett. 112:211-216(1993).
			[2] Lin Q., Ma L., Burkhart W., Spremulli L.L. J. Biol. Chem. 269:9436-9444(1994).
1F4E	PDOC00641	Eukaryotic initiation factor 4E signature	Eukaryotic translation initiation factor 4E (eIF-4E) [1] is a protein that binds to the cap structure of eukaryotic cellular mRNAs. eIF-4E recognizes and binds the 7-methylguanosine-containing (m7Gppp) cap during an early step in the initiation of protein synthesis and facilitates ribosome binding to a mRNA by inducing the unwinding of its secondary structures. eIF-4E is a conserved protein of about 25 Kd. Site directed mutagenesis
			experiments have shown [2] that a tryptophan in the central part of the sequence of human eIF-4E seems to be implicated in cap-binding. The signature pattern for eIF-4E includes this tryptophan.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [DE]-[IFY]-x(2)-F-[KR]-x(2)-[LIVM]-x-P-x-W-E-[DVA]-x(5)-G- G-[KR]-W [The first W seems to be involved in capbinding] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Thach R.E.
			[2] [2] Ueda H., Iyo H., Doi M., Inoue M., Ishida T., Morioka H., Tanaka T., Nishikawa S., Uesugi S. FEBS Lett. 280:207-210(1991).
IF5_eIF4_eIF2		eIF4-gamma/eIF5/eIF2- epsilon	Accession number: PF02020 Definition: elF4-gamma/elF5/elF2-epsilon Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1] Gathering cutoffs: 25 25 Trusted cutoffs: 26.10 26.10 Noise cutoffs: -21.50 -21.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 96060092 Reference Title: Multidomain organization of eukaryotic guanine nucleotide
			Reference Title: exchange translation initiation factor eIF-2B subunits Reference Title: revealed by analysis of conserved sequence motifs. Reference Author: Koonin EV;



_	_	_
α	$^{\prime}$	6
\sim	_	n

	_		JOSEPH TOPONOMIA
Pfam	Prosite	Full Name	Description Reference Location: Protein Sci 1995;4:1608-1617. Comment: This domain of unknown function is found at the C-terminus Comment: of several transcription initiation factors [1]. Number of members: 31
ig	PDOC00262	Immunoglobulins and major histocompatibility complex proteins signature	The basic structure of immunoglobulin (Ig) [1] molecules is a tetramer of two light chains and two heavy chains linked by disulfide bonds. There are two types of light chains: kappa and lambda, each composed of a constant domain (CL) and a variable domain (VL). There are five types of heavy chains: alpha, delta, epsilon, gamma and mu, all consisting of a variable domain (VH) and three (in alpha, delta and gamma) or four (in epsilon and mu) constant domains (CH1 to CH4). The major histocompatibility complex (MHC) molecules are made of two chains. In class I [2] the alpha chain is composed of three extracellular domains, a transmembrane region and a cytoplasmic tail. The beta chain (beta-2-microglobulin) is composed of a single extracellular domain. In class II [3], both the alpha and the beta chains are composed of two extracellular domains, a transmembrane region and a cytoplasmic tail. It is known [4,5] that the Ig constant chain domains and a single extracellular domain in each type of MHC chains are related. These homologous domains are approximately one hundred amino acids long and include a conserved intradomain disulfide bond. We developed a small pattern around the C-terminal cysteine involved in this disulfide bond which can be used to detect these category of Ig related proteins.
			Description of pattern(s) and/or profile(s) Consensus pattern [FY]-x-C-x-[VA]-x-H-Sequences known to belong to this class detected by the pattern: Ig heavy chains type Alpha C region: All, in CH2 and CH3. Ig heavy chains type Delta C region: All, in CH3. Ig heavy chains type Epsilon C region: All, in CH3. Ig heavy chains type Epsilon C region: All, in CH3 and CH4. Ig heavy chains type Gamma C region: All, in CH3 and also CH1 in some cases Ig heavy chains type Mu C region: All, in CH2, CH3 and CH4. Ig light chains type Kappa C region: In all CL except rabbit and Xenopus. Ig light chains type Lambda C region: In all CL except rabbit. MHC class I alpha chains: All, in alpha-3 domains, including in the cytomegalovirus MHC-1 homologous protein [6]. Beta-2-microglobulin: All. MHC class II alpha chains: All, in alpha-2 domains. MHC class II beta chains: All, in beta-2 domains. Other sequence(s) detected in SWISS-PROT 71. Last update May 1991 / Text revised. References [1] Gough N. Trends Biochem. Sci. 6:203-205(1981). [2] Klein J., Figueroa F. Immunol. Today 7:41-44(1986).



	*		927
Pfam .	Prosite	Full Name	Description Figueroa F., Klein J. Immunol. Today 7:78-81(1986). [4] Orr H.T., Lancet D., Robb R.J., Lopez de Castro J.A., Strominger J.L. Nature 282:266-270(1979). [5] Cushley W., Owen M.J. Immunol. Today 4:88-92(1983). [6] Beck S., Barrel B.G. Nature 331:269-272(1988).
IMPDH_C	PDOC00391	IMP dehydrogenase / GMP reductase signature	IMP dehydrogenase (EC 1.1.1.205) (IMPDH) catalyzes the rate-limiting reaction of de novo GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [1]. Inhibition of IMP dehydrogenase activity results in the cessation of DNA synthesis. As IMP dehydrogenase is associated with cell proliferation, it is a possible target for cancer chemotherapy. Mammalian and bacterial IMPDHs are tetramers of identical chains. There are two IMP dehydrogenase isozymes in humans [2]. GMP reductase (EC 1.6.6.8) catalyzes the irreversible and NADPH-dependent reductive deamination of GMP into IMP [3]. It converts nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and maintains intracellular balance of A and G nucleotides. IMP dehydrogenase and GMP reductase share many regions of sequence similarity. One of these regions is centered on a cysteine residue thought [3] to be involved in binding IMP. We have used this region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T [C is the putative IMP-binding residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update May 1991 / First entry. References [1] Collart F.R., Huberman E. J. Biol. Chem. 263:15769-15772(1988). [2] Natsumeda Y., Ohno S., Kawasaki H., Konno Y., Weber G., Suzuki K. J. Biol. Chem. 265:5292-5295(1990).
Inos-1-P_synth		Myo-inositol-1-phosphate synthase	, ,



Pfam	Prosite	Full Name	Description
			Source of seed members: Pfam-B_959 (release 4.1)
			Gathering cutoffs: 25 25
			Trusted cutoffs: 86.80 86.80
			Noise cutoffs: -219.00 -219.00
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 95066381
			Reference Title: Comparison of INO1 gene sequences and products in Candida
			Reference Title: albicans and Saccharomyces cerevisiae.
			Reference Author: Klig LS, Zobel PA, Devry CG, Losberger C;
			Reference Location: Yeast 1994;10:789-800.
			Database Reference INTERPRO; IPR002587;
			Comment: This is a family of myo-inositol-1-phosphate
			synthases.
			Comment: Inositol-1-phosphate catalyses the
			conversion of glucose-6-
			Comment: phosphate to inositol-1-phosphate, which is
			then dephosphorylated
			Comment: to inositol [1]. Inositol phosphates play an
			important role in
			Comment: signal transduction.
			Number of members: 27
IPP isomerase		Isopentenyl-diphosphate	Accession number: PF01772
		delta-isomerase	Definition: Isopentenyl-diphosphate delta-isomerase
			Author: Bashton M, Bateman A
		Alignment method of seed: Clustalw	
		Source of seed members: Pfam-B_1099 (release 4.2)	
	:	Gathering cutoffs: -88 -88	
		Trusted cutoffs: -66.70 -66.70	
		Noise cutoffs: -106.90 -106.90	
		HMM build command line: hmmbuild -F HMM SEED	
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 98409684
			Reference Title: Differential expression of two isopentenyl
			pyrophosphate Reference Title: isomerases and enhanced carotenoid
			accumulation in a
			Reference Title: unicellular chlorophyte
			Reference Author: Sun Z, Cunningham FX Jr, Gantt E;
			Reference Location: Proc Natl Acad Sci U S A
			1998;95:11482-11488.
			Reference Number: [2]
			Reference Medline: 97373600
			Reference Title: Cloning and subcellular localization of
			hamster and rat
			Reference Title: isopentenyl diphosphate dimethylallyl
			diphosphate
			Reference Title: isomerase. A PTS1 motif targets the
			enzyme to peroxisomes.
			Reference Author: Paton VG, Shackelford JE, Krisans SK;
			Reference Location: J Biol Chem 1997;272:18945-18950.
			Database Reference INTERPRO; IPR002667;
			Comment: Isopentenyl-diphosphate delta-isomerase or
			IPP isomerase EC:5.3.3.2 Comment: catalyses the interconversion of isopentenyl
			Comment: catalyses the interconversion of isopentenyl diphosphate and
			Comment: dimethylallyl diphosphate. Dimethylallyl
		1	phosphate is the initial substrate
			Comment: for the biosynthesis of carotenoids and other
			long chain isoprenoids [1].
			Number of members: 24
K-box	PDOC00302	MADS-box domain	A number of transcription factors contain a conserved domain of
	. 5000002	signature and profile	56 amino-acid
			residues, sometimes known as the MADS-box domain [E1]. They
			are listed below:
			- Serum response factor (SRF) [1], a mammalian transcription



		929
Pfam Prosite	Full Name	Description
		binds to the Serum Response Element (SRE). This is a short
		sequence of dyad symmetry located 300 bp to the 5' end of the transcription
		initiation site
		of genes such as c-fos.
		- Mammalian myocyte-specific enhancer factors 2A to 2D
		(MEF2A to MEF2D).
		These proteins are transcription factor which binds specifically
		to the
		MEF2 element present in the regulatory regions of many muscle-specific
		genes.
		- Drosophila myocyte-specific enhancer factor 2 (MEF2).
		- Yeast GRM/PRTF protein (gene MCM1) [2], a transcriptional
		regulator of
		mating-type-specific genes.
		- Yeast arginine metabolism regulation protein I (gene ARGR1 or
		ARG80).
		- Yeast transcription factor RLM1 Yeast transcription factor SMP1.
		- Arabidopsis thaliana agamous protein (AG) [3], a probable
		transcription
		factor involved in regulating genes that determines stamen
·		and carpel
		development in wild-type flowers. Mutations in the AG gene
		result in the
		replacement of the stamens by petals and the carpels by a new flower.
		- Arabidopsis thaliana homeotic proteins Apetala1 (AP1),
		Apetala3 (AP3) and
		Pistillata (PI) which act locally to specify the identity of the
		floral
		meristem and to determine sepal and petal development [4].
		- Antirrhinum majus and tobacco homeotic protein deficiens
		(DEFA) and globosa (GLO) [5]. Both proteins are transcription factors involved in the
		genetic
		control of flower development. Mutations in DEFA or GLO
		cause the
		transformation of petals into sepals and of stamina into carpels.
		- Arabidopsis thaliana putative transcription factors AGL1 to
		AGL6 [6].
		- Antirrhinum majus morphogenetic protein DEF H33 (squamosa).
		In SRF, the conserved domain has been shown [1] to be involved
		in DNA-binding
		and dimerization. We have derived a pattern that spans the
		complete length of
ĺ		the domain. The profile also spans the length of the MADS-box.
		Description of pattern(s) and/or profile(s)
		Consensus pattern R-x-[RK]-x(5)-1-x-[DNGSK]-x(3)-[KR]-x(2)-T-
		[FY]-x-[RK](3)- x(2)-[LIVM]-x-K(2)-A-x-E-[LIVM]-[STA]-x-L-x(4)-
		[LIVM]-x- [LIVM](3)-x(6)-[LIVMF]-x(2)-[FY] Sequences known to belong to this class detected by the pattern
		ALL.
		Other sequence(s) detected in SWISS-PROT NONE.
		Sequences known to belong to this class detected by the profile
		ALL.
		Other sequence(s) detected in SWISS-PROT NONE.
		Note this documentation entry is linked to both signature patterns
		and a profile. As the profile is much more sensitive than the
		patterns, you should use it if you have access to the necessary
		software tools to do so.
		Last update
		July 1999 / Pattern and text revised.
1		References
<u> </u>		
		[1] Norman C., Runswick M., Pollock R., Treisman R.



Pfam	Prosite	Full Name	Description
			Cell 55:989-1003(1988).
			[2] Passmore S., Maine G.T., Elble R., Christ C., Tye BK. J. Mol. Biol. 204:593-606(1988).
			[3] Yanofsky M., Ma H., Bowman J., Drews G., Feldmann K.A., Meyerowitz E.M. Nature 346:35-39(1990).
			[4] Goto K., Meyerowitz E.M. Genes Dev. 8:1548-1560(1994).
			[5] Troebner W., Ramirez L., Motte P., Hue I., Huijser P., Loennig WE., Saedler H., Sommer H., Schwartz-Sommer Z. EMBO J. 11:4693-4704(1992).
			[6] Ma H., Yanofsky M.F., Meyerowitz E.M. Genes Dev. 5:484-495(1991).
			[E1] http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C0014
Keratin_B2		Keratin, high sulfur B2 protein	Accession number: PF01500 Definition: Keratin, high sulfur B2 protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_706 (release 4.0) Gathering cutoffs: -17 -17 Trusted cutoffs: -1.50 -1.50 Noise cutoffs: -46.00 18.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 98201605 Reference Title: structure and hair follicle-specific expression of genes Reference Title: encoding the rat high sulfur protein B2 family. Reference Author: Hotta M, Tsuboi R, Reference Author: Adachi-Yamada T, Hotta M, Tsuboi R, Reference Location: Database Reference Comment: High sulfur proteins are cysteine-rich proteins synthesized Comment: during the differentiation of hair matrix cells, and form hair Comment: fibers in association with hair keratin intermediate filaments [1]. Comment: region containing 8 copies of a short repeat [1]. This family is Comment: also known as B2 or KAP1. Number of members: 17
ketoacyl-synt		Beta-ketoacyl synthases active site	Beta-ketoacyl-ACP synthase (EC 2.3.1.41) (KAS) [1] is the enzyme that catalyzes the condensation of malonyl-ACP with the growing fatty acid chain. It is found as a component of the following enzymatic systems: - Fatty acid synthetase (FAS), which catalyzes the formation of long-chain
			fatty acids from acetyl-CoA, malonyl-CoA and NADPH. Bacterial and plant chloroplast FAS are composed of eight separate subunits which correspond to different enzymatic activities; beta-ketoacyl synthase is one



			931
Pfam	Prosite	Full Name	Description
			of these polypeptides. Fungal FAS consists of two multifunctional
			proteins, FAS1 and
			FAS2; the beta-ketoacyl synthase domain is located in the
			C-terminal section of FAS2. Vertebrate FAS consists of a single
		·	multifunctional chain;
			the beta-ketoacyl synthase domain is located in the N-terminal
			section [2] The multifunctional 6-methysalicylic acid synthase (MSAS) from
			Penicillium
			patulum [3]. This is a multifunctional enzyme involved in the biosynthesis
			of a polyketide antibiotic and which has a KAS domain in its N-terminal
			section.
			- Polyketide antibiotic synthase enzyme systems. Polyketides are secondary
			metabolites produced by microorganisms and plants from
			simple fatty acids. KAS is one of the components involved in the biosynthesis
			of the
			Streptomyces polyketide antibiotics granatacin [4], tetracenomycin C [5]
			and erythromycin.
			- Emericella nidulans multifunctional protein Wa. Wa is involved in the
			biosynthesis of conidial green pigment. Wa is protein of 216
			contains a KAS domain.
			- Rhizobium nodulation protein nodE, which probably acts as a
			beta-ketoacyl synthase in the synthesis of the nodulation Nod factor fatty acyl
			chain.
			- Yeast mitochondrial protein CEM1.
			The condensation reaction is a two step process: the acyl
			component of an
			activated acyl primer is transferred to a cysteine residue of the enzyme and
			is then condensed with an activated malonyl donor with the
	1.		concomitant release of carbon dioxide. The sequence around the active site
			cysteine is well
			conserved and can be used as a signature pattern.
			Description of pattern(s) and/or profile(s)
			Consensus pattern G-x(4)-[LIVMFAP]-x(2)-[AGC]-C-[STA](2)-
			[STAG]-x(3)-[LIVMF] [C is the active site residue]
			Sequences known to belong to this class detected by the pattern ALL, except for bacterial and plant beta-ketoacyl synthase III
			(KAS III).
			Other sequence(s) detected in SWISS-PROT 10.
			Last update November 1997 / Text revised.
			References
			[1] Kauppinen S., Siggaard-Andersen M., von Wettstein-Knowles P.
			Carlsberg Res. Commun. 53:357-370(1988).
			[2]
			[2] Witkowski A., Rangan V.S., Randhawa Z.I., Amy C.M., Smith S.
			Eur. J. Biochem. 198:571-579(1991).
			[3]
			Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E. Eur. J. Biochem. 192:487-498(1990).
			[4]
			[4] Bibb M.J., Biro S., Motamedi H., Collins J.F., Hutchinson C.R.
			EMBO J. 8:2727-2736(1989).

Pfam	Prosite	Full Name	932
00,000	risons	un Name	Description
			[5] Sherman D.H., Malpartida F., Bibb M.J., Kieser H.M., Bibb M.J., Hopwood D.A. EMBO J. 8:2717-2725(1989).
KRAB		KRAB box	Accession number: PF01352 Definition: KRAB box Author: Bateman A Alignment method of seed: Manual Source of seed members: Bateman A Gathering cutoffs: 0 0
			Trusted cutoffs: 1.10 1.10 Noise cutoffs: -5.40 -5.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			Reference Medline: 91319563 Reference Title: Upstream from the
			Reference Title: zinc finger region of Kox 8. Reference Author: Thiesen HJ, Bellefroid E, Revelant O, Martial JA; Reference Location: Nucleic Acids Res 1991:19:3996-3996
			Reference Number: [2] Reference Medline: 97140325
			Reference Title: A novel member of the RING finger family, KRIP-1, Reference Title: associates with the KRAB-A transcriptional repressor domain
			Reference Title: of zinc finger proteins. Reference Author: Kim SS, Chen YM, O'Leary E, Witzgall R, Vidal M, Bonventre
			Reference Author: JV; Reference Location: Proc Natl Acad Sci U S A 1996;93:15299-15304.
			Reference Number: [3] Reference Medline: 96365472 Reference Title: KAP-1, a novel corepressor for the highly
	:		conserved KRAB Reference Title: repression domain. Reference Author: Friedman JR, Fredericks WJ, Jensen DE,
			Speicher DW, Huang Reference Author: XP, Neilson EG, Rauscher FJ; Reference Location: Genes Dev 1996;10:2067-2078. Database Reference INTERPRO; IPR001909;
			Database reference: PFAMB; PB036541; Comment: The KRAB domain (or Kruppel-associated box) is present in
			Comment: about a third of zinc finger proteins containing C2H2 fingers. Comment: The KRAB domain is found to be involved in protein-protein
			Comment: interactions [2,3]. Comment: The KRAB domain is generally encoded by two exons. The
			Comment: regions coded by the two exons are known as KRAB-A and Comment: KRAB-B.
			Number of members: 105
ectin_legB	PDOC00278	Legume lectins signatures	Leguminous plants synthesize sugar-binding proteins which are called legume lectins [1,2]. These lectins are generally found in the seeds. The exact
			function of legume lectins is not known but they may be involved in the attachment of nitrogen-fixing bacteria to legumes and in the protection against pathogens. Legume lectins bind calcium and
			manganese (or other transition metals).
			Legume lectins are synthesized as precursor proteins of about



		-	933
Pfam	Prosite	Full Name	Description
			230 to 260 amino
			acid residues. Some legume lectins are proteolytically
			processed to produce
			two chains: beta (which corresponds to the N-terminal) and alpha (C-terminal).
			The lectin concanavalin A (conA) from jack bean is exceptional in
			that the two
			chains are transposed and ligated (by formation of a new peptide
			bond). The
			N-terminus of mature conA thus corresponds to that of the alpha
			chain and the C-terminus to the beta chain.
			Offernings to the beta chain.
			We have developed two signature patterns specific to legume
			lectins: the first
			is located in the C-terminal section of the beta chain and
			contains a
			conserved aspartic acid residue important for the binding of calcium and
			manganese; the second one is located in the N-terminal of the
			alpha chain.
			STORING STORING
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIV] (STACLV (DEOVA (CLU D. (OT) (C.).
			Consensus pattern [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST] [D binds manganese and calcium]
			Sequences known to belong to this class detected by the pattern
		i	ALL.
			Other sequence(s) detected in SWISS-PROT 21.
			Consensus pattern [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIVF]-G-[LF]-[ST]
	i		Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT 4.
			Last update July 1999 / Patterns and text revised.
			References
			[1]
			Sharon N., Lis H.
İ			FASEB J. 4:3198-320(1990).
			r or
			[2] Lis H., Sharon N.
			Annu. Rev. Biochem. 55:33-37(1986).
ligase-CoA		CoA-ligases	Accession number: PF00549
			Definition: CoA-ligases
			Author: Bateman A
			Alignment method of seed: Clustalw Source of seed members: SCOP
			Gathering cutoffs: 25 25
1			Trusted cutoffs: 28.70 28.70
			Noise cutoffs: 14.70 14.70
			HMM build command line: hmmbuild -f HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1] Reference Medline: 94193797
			Reference Title: The crystal structure of succinyl-CoA
			synthetase from
			Reference Title: Escherichia coli at 2.5-A resolution.
:			Reference Author: Wolodko WT, Fraser ME, James MN,
			Bridger WA;
			Reference Location: J Biol Chem 1994;269:10883-10890. Database Reference: SCOP; 1scu; sf; [SCOP-USA][CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR000303;
	ļ		Database Reference PDB; 1cqi A; 132; 279;
	İ		Database Reference PDB; 1cqi D; 132; 279;
		4	Database Reference PDB; 1cqi A; 132; 279;
	1		Database Reference PDB; 1cqi D; 132; 279; Database Reference PDB: 2scu A: 132; 279:
	1		Database Reference PDB; 2scu A; 132; 279; PDB; 2scu D; 132; 279; PDB; 2scu D; 132; 279;
L	· · · · · · · · · · · · · · · · · · ·		24142435 - 1016161166 FDD, 2361 D, 132, 279,

a	3	1
~,	. ว	4

F== 2			
Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1scu A; 132; 279;
			Database Reference PDB; 1scu D; 132; 279;
			Database Reference PDB; 1cqi B; 246; 385;
			Database Reference PDB; 1cqi E; 246; 385;
			Database Reference PDB; 1cqi B; 246; 385;
			Database Reference PDB; 1cqj E; 246; 385; Database Reference PDB: 2scu B: 246; 385
	1		
			Database Reference PDB; 2scu E; 246; 385; PDB; 1scu B; 246; 388;
			Database Reference PDB; 1scu E; 246; 388;
			Database reference: PFAMB; PB039724;
İ			Database reference: PFAMB; PB041236;
			Comment: -!- This family includes the CoA ligases
			Succinyl-CoA synthetase alpha
			Comment: and beta chains, malate CoA ligase and
			ATP-citrate lyase.
			Comment: Some members of the family utilise ATP
			others use GTP.
			Number of members: 76
LIM_bind		LIM domain hinding	
LIM_DING		LIM-domain binding protein	Accession number: PF01803
		protein	Definition: LIM-domain binding protein
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_1352 (release 4.2) Gathering cutoffs: -92 -92
			Trusted cutoffs: 13.40 13.40
			Noise cutoffs: -197.90 -197.90
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 97477378
			Reference Title: Chip, a widely expressed chromosomal
	,		protein required for
			Reference Title: segmentation and activity of a remote wing
			margin enhancer
			Reference Title: in Drosophila.
			Reference Author: Morcillo P, Rosen C, Baylies MK, Dorsett
			D;
	İ		Reference Location: Genes Dev 1997;11:2729-2740.
			Reference Number: [2]
			Reference Medline: 97336071
			Reference Title: A family of LIM domain-associated cofactors confer
	İ		
			Reference Title: transcriptional synergism between LIM and Otx homeodomain
			Reference Title: proteins.
			Reference Author: Bach I, Carriere C, Ostendorff HP.
			Andersen B, Rosenfeld
			Reference Author: MG;
	ļ		Reference Location: Genes Dev 1997;11:1370-1380.
	l		Reference Number: [3]
	ļ		Reference Medline: 97078753
			Reference Title: Interactions of the LIM-domain-binding
			factor Ldb1 with LIM
			Reference Title: homeodomain proteins.
			Reference Author: Agulnick AD, Taira M, Breen JJ, Tanaka
			T, Dawid IB,
			Reference Author: Westphal H;
			Reference Location: Nature 1996;384:270-272.
			Reference Number: [4]
	-		Reference Medline: 97030257
			Reference Title: Nuclear LIM interactor, a rhombotin and LIM homeodomain
			Reference Title: interacting protein, is expressed early in neuronal
			Reference Location: Proc Natl Acad Sci U S A 1996;93:11693-11698.
			Database Reference INTERPRO; IPR002691;
			Comment: The LIM-domain binding protein, binds to
			the LIM domain LIM of
			Comment: LIM homeodomain proteins which are
·			Environmendomain proteins which are

Pfam	Prosite	Full Name	Description
	1.0010	1 MILITARITY	transcriptional regulators of
			Comment: development.
			Comment: Nuclear LIM interactor (NLI) / LIM domain-
			binding protein 1 (LDB1) Comment: Swiss:P70662 is located in the nuclei of
			neuronal cells during
			Comment: development, it is co-expressed with Isl1 in
			early motor neuron
			Comment: differentiation and has a suggested role in
			the IsI1 dependent Comment: development of motor neurons [4].
			Comment: development of motor fledions [4].
			synergistically to enhance
			Comment: transcriptional efficiency by acting as co-
			factors for LIM homeodomain
			Comment: and Otx class transcription factors both of which have essential roles
			Comment: in development [2].
			Comment: The Drosophila protein Chip Swiss:O18353
			is required for segmentation
			Comment: and activity of a remote wing margin
			enhancer [1]. Chip is a ubiquitous Comment: chromosomal factor required for normal
			expression of diverse genes at
			Comment: many stages of development [1]. It is
			suggested that Chip cooperates
			Comment: with different LIM domain proteins and other
			factors to structurally Comment: support remote enhancer-promoter
			interactions [1].
	4		Number of members: 19
Linna 2	DD0000110		T-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1
Lipase_3	PDOC00110	Lipases, serine active site	Triglyceride lipases (EC 3.1.1.3) [1] are lipolytic enzymes that hydrolyzes
		0.10	the ester bond of triglycerides. Lipases are widely distributed in
			animals,
			plants and prokaryotes. In higher vertebrates there are at least
			three tissue-
]		specific isozymes: pancreatic, hepatic, and gastric/lingual. These three types
			of lipases are closely related to each other as well as to
			lipoprotein lipase
			(EC 3.1.1.34) [2], which hydrolyzes triglycerides of chylomicrons
			and very low density lipoproteins (VLDL).
			density approxima (VESE).
			The most conserved region in all these proteins is centered
			around a serine
			residue which has been shown [3] to participate, with an histidine and an
			aspartic acid residue, to a charge relay system. Such a region is
			also present
			in lipases of prokaryotic origin and in lecithin-cholesterol
			acyltransferase (FC 2.3.1.43) (LCAT) [4], which catalyzes fatty acid transfer
			(EC 2.3.1.43) (LCAT) [4], which catalyzes fatty acid transfer between
			phosphatidylcholine and cholesterol. We have built a pattern from
			that region.
			Description of pattern(s) and/or profile(s)
	1		Conseque nothern (III) A III II III II II II II II II II II I
			Consensus pattern [LIV]-x-[LIVFY]-[LIVMST]-G-[HYWV]-S-x-G- [GSTAC] IS is the active site residue]
			Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT 35.
			Note Drosophile vitellogenine are also related to lineage IEL but
			Note Drosophila vitellogenins are also related to lipases [5], but they have lost their active site serine.
			Last update
			November 1997 / Pattern and text revised.
			References



			936
Pfam	Prosite	Full Name	Description
			[1] Chapus C., Rovery M., Sarda L., Verger R. Biochimie 70:1223-1234(1988).
			[2] Persson B., Bengtsson-Olivecrona G., Enerback S., Olivecrona T., Joernvall H. Eur. J. Biochem. 179:39-45(1989).
			[3] Blow D. Nature 343:694-695(1990).
			[4] McLean J., Fielding C., Drayna D., Dieplinger H., Baer B., Kohr W., Henzel W., Lawn R. Proc. Natl. Acad. Sci. U.S.A. 83:2335-2339(1986).
			[5] Baker M.E. Biochem. J. 255:1057-1060(1988).
Lipase_GDSL	PDOC00842	Lipolytic enzymes "G-D- S-L" family, serine active site	Recently [1], a family of lipolytic enzymes has been characterized. This family currently consist of the following proteins:
			- Aeromonas hydrophila lipase/phosphatidylcholine-sterol acyltransferase Xenorhabdus luminescens lipase 1 Vibrio mimicus arylesterase Escherichia coli acyl-coA thioesterase I (gene tesA) Vibrio parahaemolyticus thermolabile hemolysin/atypical phospholipase Rabbit phospholipase AdRab-B, an intestinal brush border protein with esterase and phospholipase A/lysophospholipase activity that could be involved in the uptake of dietary lipids. AdRab-B contains four repeats of about 320 amino acids Arabidopsis thaliana and Brassic napus anther-specific prolinerich protein APG A Pseudomonas putida hypothetical protein in trpE-trpG intergenic region. A serine has been identified a part of the active site in the Aeromonas, Vibrio mimicus and Escherichia coli enzymes. It is located in a conserved sequence motif that can be used as a signature pattern for these proteins.
			Description of pattern(s) and/or profile(s) Consensus pattern [LIVMFYAG](4)-G-D-S-[LIVM]-x(1,2)-[TAG]-G [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this pattern will pick up two of the four repeats in AdRab-B, the first one is not detected as its sequence has diverged in the region of the putative active site residue. The last one is also not detected because it is slightly divergent at the end of the pattern. Expert(s) to contact by email
			Upton C. upton@sol.uvic.ca Buckley J.T. tbuckley@sol.uvic.ca
			Last update November 1995 / First entry.

			937
Pfam	Prosite	Full Name	Description
			References
			[[1] Upton C., Buckley J.T.
			Trends Biochem. Sci. 20:178-179(1995).
Lipoprotein_1	PDOC00013	Prokaryotic membrane lipoprotein lipid	In prokaryotes, membrane lipoproteins are synthesized with a precursor signal
		attachment site	peptide, which is cleaved by a specific lipoprotein signal peptidase (signal
			peptidase II). The peptidase recognizes a conserved sequence and cuts upstream
	II.		of a cysteine residue to which a glyceride-fatty acid lipid is attached [1].
			Some of the proteins known to undergo such processing
			currently include (for
			recent listings see [1,2,3]):
			- Major outer membrane lipoprotein (murein-lipoproteins) (gene
			- Escherichia coli lipoprotein-28 (gene nlpA).
			- Escherichia coli lipoprotein-34 (gene nlpB).
			- Escherichia coli lipoprotein nIpC.
			Escherichia coli lipoprotein nlpD. Escherichia coli osmotically inducible lipoprotein B (gene
			osmB).
			- Escherichia coli osmotically inducible lipoprotein E (gene osmE).
			- Escherichia coli peptidoglycan-associated lipoprotein (gene
			- Escherichia coli rare lipoproteins A and B (genes rplA and rplB).
			- Escherichia coli copper homeostasis protein cutF (or nlpE).
			- Escherichia coli plasmids traT proteins.
			Escherichia coli Col plasmids lysis proteins. A number of Bacillus beta-lactamases.
			- Bacillus subtilis periplasmic oligopeptide-binding protein (gene
			oppA).
			- Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB).
			- Borrelia hermsii variable major protein 21 (gene vmp21) and 7
			(gene vmp7) Chlamydia trachomatis outer membrane protein 3 (gene omp3).
			- Fibrobacter succinogenes endoglucanase cel-3.
			- Haemophilus influenzae proteins Pal and Pcp.
			- Klebsiella pullulunase (gene pulA).
			Klebsiella pullulunase secretion protein pulS. Mycoplasma hyorhinis protein p37.
			- Mycoplasma hyorhinis protein p37. - Mycoplasma hyorhinis variant surface antigens A, B, and C
			(genes vlpABC).
	i		- Neisseria outer membrane protein H.8.
			- Pseudomonas aeruginosa lipopeptide (gene lppL).
			Pseudomonas solanacearum endoglucanase egl. Rhodopseudomonas viridis reaction center cytochrome subunit
			(gene cytC).
	İ		- Rickettsia 17 Kd antigen.
			- Shigella flexneri invasion plasmid proteins mxiJ and mxiM Streptococcus pneumoniae oligopeptide transport protein A
			(gene amiA).
			- Treponema pallidium 34 Kd antigen.
			- Treponema pallidium membrane protein A (gene tmpA).
			- Vibrio harveyi chitobiase (gene chb). - Yersinia virulence plasmid protein yscJ.
			' '
			- Halocyanin from Natrobacterium pharaonis [4], a membrane
			associated copper- binding protein. This is the first archaebacterial protein
			known to be
			modified in such a fashion).
			From the precursor sequences of all these proteins, we derived a consensus
			pattern and a set of rules to identify this type of post-
			translational
1			modification.

			938
Pfam	Prosite	Full Name	Description
Lipoprotein_2	PDOC00013		Description Description of pattern(s) and/or profile(s) Consensus pattern (DERK)(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be. Last update November 1995 / Pattern and text revised. References [1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990). [2] Kilein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:531-534(1989). [3] von Heijne G. Protein Eng. 2:531-534(1989). [4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994). In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptidae, which is cleaved by a specific lipoprotein signal peptidae (signal peptidase) II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]): - Major outer membrane lipoprotein (murein-lipoproteins) (gene lipp) Escherichia coli lipoprotein-28 (gene nipA) Escherichia coli lipoprotein-34 (gene nipB) Escherichia coli lipoprotein nipD Escherichia coli lipoprotein nipD Escherichia coli lipoprotein nipD Escherichia coli lipoprotein nipD Escherichia coli lipoprotein nipD Escherichia coli lipoprotein nipD Escherichia coli lipoprotein nipD Escherichia coli lipoprotein nipD Escherichia coli lipoprotein nipD Escherichia coli lipoprotein nipD Escherichia coli peptidoglycan-associated lipoprotein (gene pal) Escherichia coli peptidoglycan-associated lipoprotein (sepen pal)
			- Escherichia coli lipoprotein nlpD Escherichia coli osmotically inducible lipoprotein B (gene osmB) Escherichia coli osmotically inducible lipoprotein E (gene osmE) Escherichia coli peptidoglycan-associated lipoprotein (gene pal) Escherichia coli rare lipoproteins A and B (genes rpIA and rpIB) Escherichia coli copper homeostasis protein cutF (or nlpE) Escherichia coli plasmids traT proteins Escherichia coli Col plasmids lysis proteins.



		9	39
Pfam	Prosite	Full Name	Description (copes vinABC)
			(genes vlpABC). Neisseria outer membrane protein H.8. Pseudomonas aeruginosa lipopeptide (gene lppL). Pseudomonas solanacearum endoglucanase egl. Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC). Rickettsia 17 Kd antigen. Shigella flexneri invasion plasmid proteins mxiJ and mxiM. Streptococcus pneumoniae oligopeptide transport protein A (gene amiA). Treponema pallidium 34 Kd antigen. Treponema pallidium membrane protein A (gene tmpA). Vibrio harveyi chitobiase (gene chb). Yersinia virulence plasmid protein yscJ. Halocyanin from Natrobacterium pharaonis [4], a membrane associated copperbinding protein. This is the first archaebacterial protein known to be modified in such a fashion). From the precursor sequences of all these proteins, we derived a consensus pattern and a set of rules to identify this type of post-translational modification.
			Description of pattern(s) and/or profile(s) Consensus pattern {DERK}(6)-[LIVMFWSTAG](2)- [LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be. Last update November 1995 / Pattern and text revised. References [1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990). [2] Klein P., Somorjai R.L., Lau P.C.K.
			Protein Eng. 2:15-20(1988). [3] von Heijne G. Protein Eng. 2:531-534(1989). [4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M.
Lipoprotein_5	PDOC00013	Prokaryotic membrane lipoprotein lipid attachment site	In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]):
			- Major outer membrane lipoprotein (murein-lipoproteins) (gene

_	
Ī	
£	
I	
Ē	
<u>ا</u>	
Ø	
e Li	
e Hi	
Ø	

Pfam	Prosite	EuthNe	940
I IANI	PTOSKE	Full Name	Description Ipp).
			- Escherichia coli lipoprotein-28 (gene nlpA).
			- Escherichia coli lipoprotein-34 (gene nlpB).
			- Escherichia coli lipoprotein nlpC.
			- Escherichia coli lipoprotein nlpD.
			- Escherichia coli osmotically inducible lipoprotein B (gene
			osmB).
			- Escherichia coli osmotically inducible lipoprotein E (gene osmE).
			- Escherichia coli peptidoglycan-associated lipoprotein (gene
			(pai).
			- Escherichia coli rare lipoproteins A and B (genes rplA and rplB).
			- Escherichia coli copper homeostasis protein cutF (or plpE)
			- Escherichia coli plasmids traT proteins Escherichia coli Col plasmids lysis proteins.
			- A number of Bacillus beta-lactamases.
			- Bacillus subtilis periplasmic oligopeptide-binding protein (gene
1			(oppA).
			- Borrelia burgdorferi outer surface proteins A and B (genes ospA
			(and ospB).
			- Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7).
			- Chlamydia trachomatis outer membrane protein 3 (gene omp3)
	ŀ		- Fibrobacter succinogenes endoglucanase cel-3.
			- Haemophilus influenzae proteins Pal and Pcp
			- Klebsiella pullulunase (gene pulA).
			 Klebsiella pullulunase secretion protein pulS. Mycoplasma hyorhinis protein p37.
			- Mycoplasma hyorhinis variant surface antigens A, B, and C
			(genes vipABC).
1			- Neisseria outer membrane protein H.8.
			- Pseudomonas aeruginosa lipopeptide (gene lppl)
			- Pseudomonas solanacearum endoglucanase egl.
			 Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC).
			- Rickettsia 17 Kd antigen.
			- Shigella flexneri invasion plasmid proteins mxi.l and mxiM
	1		- Streptococcus pneumoniae oligopeptide transport protein A
İ			(gene amiA).
			- Treponema pallidium 34 Kd antigen Treponema pallidium membrane protein A (gene tmpA).
			- Vibrio harveyi chitobiase (gene chb).
			- Yersinia virulence plasmid protein yscJ.
			- Halocyanin from Natrobacterium pharaonis [4], a membrane
			associated copper- binding protein. This is the first archaebacterial protein
	1		known to be
			modified in such a fashion).
	-		From the precursor sequences of all these proteins, we derived
	į		a consensus
			pattern and a set of rules to identify this type of post- translational
			modification.
			Description of pattern (a) and (c)
Ĭ			Description of pattern(s) and/or profile(s)
			Consensus pattern {DERK}(6)-[LIVMFWSTAG](2)-
			[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site]
			Additional rules: 1) The cysteine must be between positions 15
			and 35 of the sequence in consideration, 2) There must be at
			least one Lys or one Arg in the first seven positions of the sequence.
ŀ	İ		
			Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT some 100
			prokaryotic proteins. Some of them are not membrane
			lipoproteins, but at least half of them could be.
			Last update
			November 1995 / Pattern and text revised.

		le usi	December
Pfam	Prosite	Full Name	Description [1]
			Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).
			[2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).
			[3] von Heijne G. Protein Eng. 2:531-534(1989).
			[4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).
			3. Biol. Offerit. 203. (4935-14943).
Luteo_Vpg		Luteovirus putative VPg genome linked protein	Accession number: PF01659 Definition: Luteovirus putative VPg genome linked protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_970 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 191.70 191.70 Noise cutoffs: -47.90 -47.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 94120742 Reference Title: Soybean dwarf luteovirus contains the third variant genome Reference Title: type in the luteovirus group. Reference Author: Rathjen JP, Karageorgos LE, Habili N, Waterhouse PM, Symons Reference Location: Virology 1994;198:671-679. Database Reference Comment: This family consists of several putative genome linked proteins. Comment: The genomic RNA of luteoviruses are linked to virally encoded genome Comment: proteins (VPg). Open reading frame 4 is thought to encode the VPg Comment: in Soybean dwarf luteovirus [1]. Comment: Luteoviruses have isometric capsids that contain a positive stand Comment: ssRNA genome, they have no DNA stage during their replication.
		NATIL de la circ	Number of members: 32
MATH		MATH domain	Accession number: PF00917 Definition: MATH domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1602 (release 3.0) Gathering cutoffs: 17 0 Trusted cutoffs: 17.90 0.20 Noise cutoffs: 11.80 11.80 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmbcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 96334294 Reference Title: TRAF proteins and meprins share a conserved domain. Reference Author: Uren AG, Vaux DL; Reference Number: [2] Reference Number: [2] Reference Title: Crystallographic analysis of CD40 recognition and signaling Reference Title: McWhirter SM, Pullen SS, Holton JM, Crute JJ, Kehry MR, Reference Author: Alber T;



Dfam	Droots		942
Pfam	Prosite	Full Name	Description Reference Location: Proc Natl Acad Sci U S A 1999;96:8408-
			8413. Reference Number: [3]
			Reference Number: [3] Reference Medline: 99069615
			Reference Title: Comparison of the complete protein sets of
i			worm and yeast: Reference Title: orthology and divergence.
			Reference Title: orthology and divergence. Reference Author: Chervitz SA, Aravind L, Sherlock G, Ball
	l		CA, Koonin EV,
	į.		Reference Author: Dwight SS, Harris MA, Dolinski K, Mohr S,
			Smith T, Weng S,
			Reference Author: Cherry JM, Botstein D; Reference Location: Science 1998;282:2022-2028.
			Database Reference: SCOP; 1qsc; fa; [SCOP-USA][CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR002083;
			Database Reference PDB; 1qsc A; 357; 498;
			Database Reference PDB; 1qsc B; 357; 498;
			Database Reference PDB; 1qsc C; 357; 498; Database reference: PFAMB; PB018448;
			Database reference: PFAMB; PB040690;
			Database reference: PFAMB; PB041198;
			Comment: This motif has been called the Meprin And
			TRAF-Homology
			Comment: (MATH) domain. This domain is hugely expanded in the nematode
			Comment: C. elegans [3].
			Number of members: 212
MCT	;	Monocarboxylate	Accession number: PF01587
		transporter	Definition: Monocarboxylate transporter Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_483 (release 4.1)
			Gathering cutoffs: 25 25
			Trusted cutoffs: 322.90 322.90
			Noise cutoffs: -38.20 -38.20 HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 98087501
			Reference Title: Cloning and sequencing of four new
			mammalian Reference Title: monocarboxylate transporter (MCT)
			homologues confirms the
			Reference Title: existence of a transporter family with an
			ancient past.
			Reference Author: Price NT, Jackson VN, Halestrap AP, Biochem J 1998;329:321-328.
			Database Reference INTERPRO; IPR002897;
			Comment: This domain consists of the transmembrane
			region of the monocarboxylate
			Comment: transporters. Monocarboxylate transporters
			(MTC) are transmembrane Comment: glycoproteins with 10-12 predicted
			transmembrane regions.
			Comment: They catalyse the proton linked transport of
			lactic acid,
			Comment: pyruvate and ketone bodies across the
			plasma membrane [1]. Number of members: 33
			Tannos of moniporo.
Methionine_synt		Methionine synthase,	Accession number: PF01717
		vitamin-B12 independent	1
			independent Author: Rechton M. Retemon A
	l		Author: Bashton M, Bateman A Alignment method of seed: Clustalw
			Source of seed members: Pfam-B 1909 (release 4.1)
			Gathering cutoffs: -155.0 -155.0
			Trusted cutoffs: -155.00 -155.00
			Noise cutoffs: -170.00 -170.00
			HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Transcription [1]



	1 - .	I=	943
Pfam	Prosite	Full Name	Description Reference Medline: 98301657
			Reference Title: The specific features of methionine
			biosynthesis and Reference Title: metabolism in plants.
			Reference Author: Ravanel S, Gakiere B, Job D, Douce R;
			Reference Location: Proc Natl Acad Sci U S A 1998;95:7805-
			7812.
			Database Reference INTERPRO; IPR002629;
			Database reference: PFAMB; PB041617; Comment: PFAMB; PB041617; This is a family of vitamin-B12 independent
			Comment: This is a family of vitamin-B12 independent methionine synthases
			Comment: or 5-methyltetrahydropteroyltriglutamate
			homocysteine
			Comment: methyltransferases, EC:2.1.1.14 from
			bacteria and plants.
			Comment: Plants are the only higher eukaryotes that
			have the required enzymes Comment: for methionine synthesis [1].
		7	Comment: This enzyme catalyses the last step in the
			production of methionine
			Comment: by transferring a methyl group from 5-
			methyltetrahydrofolate to
			Comment: homocysteine [1].
			Comment: The aligned region makes up the carboxy region of the approximately
			Comment: 750 amino acid protein except in some
			hypothetical archaeal proteins
			Comment: present in the family, where this region
			corresponds to the
			Comment: entire length.
			Number of members: 28
Methyltransf 2		O-methyltransferase	Accession number: PF00891
Wictifyitianisi_E		O mearymanororade	Definition: O-methyltransferase
			Previous Pfam IDs: Methyltransf;
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_152 (release 3.0) Gathering cutoffs: -53 -53
		1	Trusted cutoffs: -22.00 -22.00
			Noise cutoffs: -84.60 -84.60
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 93167811 Reference Title: Purification of a 40-kilodalton
			methyltransferase active in
			Reference Title: the aflatoxin biosynthetic pathway.
	1		1 to or or or or or or or or or or or or or
İ	1		Reference Author: Keller NP, Dischinger HC, Bhatnagar D,
			Reference Author: Keller NP, Dischinger HC, Bhatnagar D, Cleveland TE, Ullah
			Reference Author: Keller NP, Dischinger HC, Bhatnagar D, Cleveland TE, Ullah Reference Author: AH;
			Reference Author: Keller NP, Dischinger HC, Bhatnagar D, Cleveland TE, Ullah Reference Author: AH; Appl Environ Microbiol 1993;59:479-484.
			Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Keller NP, Dischinger HC, Bhatnagar D, AH; Appl Environ Microbiol 1993;59:479-484. INTERPRO; IPR001077;
			Reference Author: Keller NP, Dischinger HC, Bhatnagar D, Cleveland TE, Ullah Reference Author: AH; Appl Environ Microbiol 1993;59:479-484.
			Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: INTERPRO; IPR001077; This family includes a range of Omethyltransferases. These Comment: enzymes utilise S-adenosyl methionine.
			Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of O- methyltransferases. These Keller NP, Dischinger HC, Bhatnagar D, AH; Appl Environ Microbiol 1993;59:479-484. INTERPRO; IPR001077; This family includes a range of O-
		O mathe the sect	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of Ommethyltransferases. These Comment: Number of members: Keller NP, Dischinger HC, Bhatnagar D, AH; Appl Environ Microbiol 1993;59:479-484. INTERPRO; IPR001077; This family includes a range of Ommethyltransferases. These enzymes utilise S-adenosyl methionine.
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: methyltransferases. These Comment: Number of members: Keller NP, Dischinger HC, Bhatnagar D, AH; Appl Environ Microbiol 1993;59:479-484. INTERPRO; IPR001077; This family includes a range of O- methyltransferases. These enzymes utilise S-adenosyl methionine. Accession number: PF01596
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of Ommethyltransferases. These Comment: Number of members: Keller NP, Dischinger HC, Bhatnagar D, AH; Appl Environ Microbiol 1993;59:479-484. INTERPRO; IPR001077; This family includes a range of Ommethyltransferases. These enzymes utilise S-adenosyl methionine.
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: methyltransferases. These Comment: Number of members: Keller NP, Dischinger HC, Bhatnagar D, AH; Appl Environ Microbiol 1993;59:479-484. INTERPRO; IPR001077; This family includes a range of O-methyltransferase enzymes utilise S-adenosyl methionine. Accession number: PF01596 Definition: O-methyltransferase
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of O-methyltransferases. These Comment: Number of members: Accession number: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PKeller NP, Dischinger HC, Bhatnagar D, AH; AH; APpl Environ Microbiol 1993;59:479-484. INTERPRO; IPR001077; This family includes a range of O-methyltransferase enzymes utilise S-adenosyl methionine. PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1)
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference INTERPRO; IPR001077; Comment: This family includes a range of Omethyltransferases. These Comment: Number of members: Accession number: Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PKeller NP, Dischinger HC, Bhatnagar D, AH; Appl Environ Microbiol 1993;59:479-484. INTERPRO; IPR001077; This family includes a range of Omethyltransferase and this control of the property of the prope
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of O-methyltransferases. These Comment: Number of members: Accession number: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pf01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pf01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1) Gathering cutoffs: -86 -86 Trusted cutoffs: -81.80 -81.80
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of Omethyltransferases. These Comment: Number of members: Accession number: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1) Gathering cutoffs: -81.80 -81.80 Noise cutoffs: -91.00 -91.00
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of Omethyltransferases. These Comment: Number of members: Accession number: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1) Gathering cutoffs: -81.80 -81.80 Noise cutoffs: -91.00 -91.00 HMM build command line: hmmbuild -F HMM SEED
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Aeference Location: Database Reference Comment: This family includes a range of O-methyltransferases. These Comment: Number of members: Accession number: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1) Gathering cutoffs: -81.80 -81.80 Noise cutoffs: -91.00 -91.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of O-methyltransferases. These Comment: Number of members: Accession number: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1) Gathering cutoffs: -81.80 -81.80 Noise cutoffs: -91.00 -91.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of O-methyltransferases. These Comment: Number of members: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1) Gathering cutoffs: -81.80 -81.80 Noise cutoffs: -91.00 -91.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of Omethyltransferases. These Comment: Number of members: Accession number: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1) Gathering cutoffs: -81.80 -81.80 Noise cutoffs: -91.00 -91.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: Reference Medline: Reference Title: Two multifunctional peptide synthetases and an
Methyltransf_3		O-methyltransferase	Reference Author: Cleveland TE, Ullah Reference Author: Reference Location: Database Reference Comment: This family includes a range of Omethyltransferases. These Comment: Number of members: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1) Gathering cutoffs: -86 -86 Trusted cutoffs: -81.80 -81.80 Noise cutoffs: -91.00 -91.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: Reference Title: Two multifunctional peptide synthetases

Pfam	Prosite	Full Name	Description	
L. C. iosao	I diri vaine	Reference Title:	DNA-binding antibiotic and antitumour	
			agent saframycin Mx1	
			, 0	from Myxococcus xanthus.
			Reference Author:	Pospiech A, Bietenhader J, Schupp T;
			Reference Location:	Microbiology 1996;142:741-746.
			Database Reference: PDBSUM]	SCOP; 1vid; fa; [SCOP-USA][CATH-
			Database Reference	INTERPRO; IPR002935;
			Database Reference	PDB; 1vid ; 13; 186;
			Database reference:	PFAMB; PB040269;
			Comment:	Members of this family are O-
			methyltransferases. T	he family
				includes catechol o-methyltransferase
			Swiss:P21964, caffeo	
				O-methyltransferase Swiss:Q43095 and a
			family of bacterial	O and the discount of a constant of the state of the stat
			Comment:	O-methyltransferases that may be involved
				production [1].
			Number of members:	39
			inditabel of members.	39
MMR HSR1		GTPase of unknown	Accession number:	PF01926
_		function	Definition: G1	Pase of unknown function
				right A, Ouzounis C, Bateman A
			Alignment method of s	
			Source of seed memb	
			Gathering cutoffs:	-21 -21
				-20.70 -20.70
				31.60 -31.60
				line: hmmbuild HMM SEED
			Reference Number:	line: hmmcalibrateseed 0 HMM
			Reference Medline:	[1] 94235953
				Structure and evolution of a member of a
			new subfamily of	circulate and evolution of a member of a
				GTP-binding proteins mapping to the
			human MHC class I	arr amang protonic mapping to are
				region.
			Reference Author:	Vernet C, Ribouchon MT, Chimini
			GPontarotti P;	
			Reference Location:	Mamm Genome 1994;5:100-105.
			Database Reference	INTERPRO; IPR002917;
			Database reference:	PFAMB; PB000471;
			Database reference:	PFAMB; PB002171;
			Database reference:	PFAMB; PB015790;
			Number of members:	67
MoaC		MoaC family	Accession number:	PF01967
				oaC family
			Author: En	right A, Ouzounis C, Bateman A
			Alignment method of s	seed: Clustalw
			Source of seed memb	· ·
			Gathering cutoffs:	25 25
			r	73.00 73.00
				93.90 -93.90
				line: hmmbuild -F HMM SEED
				line: hmmcalibrateseed 0 HMM
			Reference Number: Reference Medline:	[1]
			Reference Mediine:	99337076 Characterization of a molybdenum cofactor
			biosynthetic gene	Characterization of a molypuerium colactor
			Reference Title:	cluster in Rhodobacter capsulatus which is
			specific for the	STATES III I III SOUDANTO OAPSUICIUS WINOIT IS
			Reference Title:	biogenesis of dimethylsulfoxide reductase.
			Reference Author:	Solomon PS, Shaw AL, Lane I, Hanson
			GR, Palmer T, McEwa	
			Reference Author:	AG;
			Reference Location:	Microbiology 1999;145:1421-1429.
			Database Reference	INTERPRO; IPR002820;
			Comment:	Members of this family are involved in
			molybdenum	
				cofactor biosynthesis. However their
	1	1	moloculor	
			molecular Comment:	function is not known.



Pfam	Prosite	Full Name	Description 24
			Number of members: 24
Morbilli_P		Morbillivirus RNA polymerase alpha subunit	Accession number: PF01647 Definition: Morbillivirus RNA polymerase alpha subunit Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_903 (release 4.1) Gathering cutoffs: -74 -74 Trusted cutoffs: 22.90 22.90 Noise cutoffs: -171.70 -171.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 92341068 Sequence analysis of the genes encoding the nucleocapsid Reference Title: distemper virus, Reference Title: and editing of the P gene transcript. Reference Author: Blixenkrone-Moller M, Sharma B, Varsany TM, Hu A, Norrby Reference Location: J. Gen. Virol. 1992;73:885-893. INTERPRO; IPR002581; PFAMB; PB002389; This family consists of morbillivirus RNA polymerase alpha subunit comment: and non structural protein V. The P gene of morbillivirus is cotranscriptionally edited leading to the N-terminal Comment: and a cysteine rich region in the V fusion protein which has been Comment: shown to bind zinc [see Virology 3rd edition volume 1, chapter 40, Comment: pages 1182-1184]. Comment: pages 1182-1184]. Comment: pages 1182-1184]. Comment: pages may yoviridae family, members include
Myc_N_term		Myc amino-terminal region	measles virus and phocine Comment: distemper virus. Number of members: 52 Accession number: PF01056 Definition: Myc amino-terminal region Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_387 (release 3.0) Gathering cutoffs: -109 -109 Trusted cutoffs: -81.20 -81.20 Noise cutoffs: -137.40 -137.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 98280742 Reference Medline: 98280742 Reference Title: The molecular role of Myc in growth and transformation: Reference Title: recent discoveries lead to new insights. Reference Author: Facchini LM, Penn LZ; Reference Number: [2] Reference Medline: 97318600 Reference Title: Myc target genes. Reference Author: Grandori C, Eisenman RN; Reference Location: Database Reference Comment: Trends Biochem Sci 1997;22:177-181. INTERPRO; IPR002418; The myc family belongs to the basic helix-loop-helix leucine zipper Comment: class of transcription factors, see HLH. M forms a Comment: heterodimer with Max, and this complex regulates cell growth through

			946
Pfam	Prosite	Full Name	Description
			Comment: direct activation of genes involved in cell
			replication [2].
			Number of members: 56
Myosin_tail		Myosin tail	Accession number: PF01576
			Definition: Myosin tail
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_356 (release 4.1)
			Gathering cutoffs: 19 19
			Trusted cutoffs: 23.30 23.30
			Noise cutoffs: 15.10 15.10
			HMM build command line: hmmbuild -f HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 87060988
		ŀ	Reference Title: Complete nucleotide and encoded amino
		1	acid sequence of a
			Reference Title: mammalian myosin heavy chain gene.
			Evidence against
	1		Reference Title: intron-dependent evolution of the rod.
			Reference Author: Strehler EE, Strehler-page M-A, Perriard
			JC, Periasamy M,
	1		Reference Author: Nadal-ginard B;
			Reference Location: J MOL BIOL 1986;190:291-317.
			Database Reference INTERPRO; IPR002928;
			Comment: The myosin molecule is a multi-subunit
			complex made up
			Comment: of two heavy chains and four light chains it is
			a fundamental contractile
			Comment: protein found in all eukaryote cell types [1].
			Comment: This family consists of the coiled-coil myosin
			heavy chain tail region.
			Comment: The coiled-coil is composed of the tail from
			two molecules of myosin.
			Comment: These can then assemble into the
			macromolecular thick filament [1].
			Comment: The coiled-coil region provides the structural
			backbone the thick
			Comment: filament [1]. Number of members: 182
la Ala suma	BDOC00681	Codiumiologica	It has been shown [1] that integral membrane proteins that
la_Ala <u>-</u> symp	PDOC00681	Sodium:alanine	It has been shown [1] that integral membrane proteins that
		symporter family	mediate the intake
		signature	of a wide variety of molecules with the concomitant uptake of
			sodium ions (sodium symporters) can be grouped, on the basis of sequence
			and functional
	•		similarities into a number of distinct families. One of these families is
			known as the sodium:alanine symporter family (SAF) and
		į	currently consists of
			the following proteins:
			- Thermophilic bacterium PS-3 alanine carrier protein (ACP).
			ACP can use both
	1		sodium and hydrogen as a symport ion.
			- Alteromonas haloplanktis D-alanine/glycine permease (gene
			dagA).
		,	- Bacillus subtilis alsT.
			- Hypothetical protein yaaJ from Escherichia coli and
			HI0183, the
			corresponding Haemophilus influenzae protein Haemophilus influenzae hypothetical protein HI0883.
			These integral membrane proteins are predicted to comprise a
		1	least eight
			membrane spanning domains. As a signature pattern we
			selected a highly
	1	ł	
	1		conserved region which is located in the N-terminal section and
			which includes



Dr	Ina-		947 15
Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s)
			Consensus pattern G-G-x-[GA](2)-[LIVM]-F-W-M-W-[LIVM]-x- [STAV]-[LIVMFA](2)-G
			Sequences known to belong to this class detected by ôhe pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Last update November 1997 / Pattern and text revised.
			References [1]
			Reizer J., Reizer A., Saier M.H. Jr. Biochim. Biophys. Acta 1197:133-136(1994).
N- 0- E-		0 - 1: / 1 - :	, i
Na_Ca_Ex		Sodium/calcium exchanger protein	Accession number: PF01699 Definition: Sodium/calcium exchanger protein
			Author: Bashton M, Bateman A Alignment method of seed: Clustalw
l			Source of seed members: Pfam-B_1680 (release 4.1)
İ			Gathering cutoffs: 3 3 Trusted cutoffs: 3.40 3.40
			Noise cutoffs: 1.20 1.20
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			Reference Number: [1] Reference Medline: 96394663
			Reference Title: Cloning of a third mammalian Na+-Ca2+
			exchanger, NCX3. Reference Author: Nicoll DA, Quednau BD, Qui Z, Xia YR,
			Lusis AJ, Philipson
			Reference Author: KD;
			Reference Location: J Biol Chem 1996;271:24914-24921. Reference Number: [2]
	:		Reference Medline: 91047958
			Reference Title: Molecular cloning and functional expression
			of the cardiac Reference Title: sarcolemmal Na(+)-Ca2+ exchanger.
			Reference Title: sarcolemmal Na(+)-Ca2+ exchanger. Reference Author: Nicoll DA, Longoni S, Philipson KD;
			Reference Location: Science 1990;250:562-565.
			Database Reference INTERPRO; IPR002613;
			Database reference: PFAMB; PB002768; Database reference: PFAMB; PB040773;
			Database reference: PFAMB; PB041540;
			Comment: This is a family of sodium/calcium
			exchanger integral membrane Comment: proteins. This family covers the integral
			membrane regions of
			Comment: the proteins. Sodium/calcium exchangers
			regulate intracellular Ca2+ Comment: concentrations in many cells; cardiac
			myocytes, epithelial cells,
			Comment: neurons retinal rod photoreceptors and smooth muscle cells [2].
			Comment: Ca2+ is moved into or out of the cytosol
			depending on Na+ concentration
			Comment: [2]. In humans and rats there are 3 isoforms; NCX1 NCX2 and NCX3 [1]
			Comment: see Swiss:Q01728, Swiss:P48768 and
			Swiss:P70549 respectively. Number of members: 105
Na Galacto symp	PDOC00680	Sodium:galactoside	It has been shown [1] that integral membrane proteins that
Na_Galacto_symp	F DOC00000	symporter family	mediate the intake
		signature	of a wide variety of molecules with the concomitant uptake of sodium ions
			(sodium symporters) can be grouped, on the basis of sequence
			and functional
			similarities into a number of distinct families. One of these families is
			known as the sodium:galactoside symporter family (SGF) and
			currently consists
			of the following proteins:
	l		<u> </u>



	12		948
Pfam	Prosite	Full Name	Description - The melibiose carrier (gene melB) from a variety of
			enterobacteria. This
			protein is responsible for melibiose transport and is capable
			of using
			hydrogen, sodium, and lithium cations as coupling cations for
			cotransport.
			- The lactose permease from Lactobacillus (gene lacS or lacY).
			This protein
			is responsible for the transport of beta-galactosides into the cell, with
			the concomitant export of a proton. It consists of two
			domains; a N-
			terminal SGF domain and a C-terminal domain that resembles
			that of enzyme
			IIA of the PEP:sugar phosphotransferase system.
			- The raffinose permease from Pediococcus pentosaceus. It also
			consists of a
			N-terminal SGF domain and a C-terminal IIA domain The glucuronide carrier (gene gusB or uidP) from Escherichia
			coli.
			- The xylose transporter (gene xylP) from Lactobacillus pentosus.
			- Escherichia coli hypothetical protein yagG.
			- Escherichia coli hypothetical protein yicJ.
			- Escherichia coli hypothetical protein yihO.
			- Escherichia coli hypothetical protein yihP.
	1		- Bacillus subtilis hypothetical protein yimB.
			- Bacillus subtilis hypothetical protein ynaJ.
			Like sugar transport proteins, these integral membrane proteins
			are predicted
			to comprise twelve membrane spanning domains. As a
			signature pattern we
			selected a highly conserved region which is located in a
			cytoplasmic loop
1	1		between the second and third transmembrane regions. This region starts with
I			a conserved aspartate which has been shown [2], in melB, to be
I			important for
1			the activity of the protein.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [DG]-x(3)-G-x(3)-[DN]-x(6,8)-[GA]-[KRHQ]-
			[FSA]-[KR]-[PT]- [FYW]-[LIVMWQ]-[LIV]-x-[GAFV]-[GSTA]
			Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE. Last update
			July 1999 / Pattern and text revised.
			References
			[1]
			Reizer J., Reizer A., Saier M.H. Jr.
			Biochim. Biophys. Acta 1197:133-136(1994).
			[2]
			2 Pourcher T., Deckert M., Bassilana M., Leblanc G.
			Biochem. Biophys. Res. Commun. 178:1176-1181(1991).
		No. ///. ATD O	This demain is excellent to the codium and notaceium ATD
Na_K_ATPase_C		Na+/K+ ATPase C- terminus	This domain is specific to the sodium and potassium ATPases (Na K-ATPase).
	1	Commus	The sodium pump (Na+,K+ ATPase), located in the plasma
			membrane of all animal cells [1], is an heterotrimer of a catalytic
			subunit (alpha chain), a glycoprotein subunit of about 34 Kd (beta
			chain) and a small hydrophobic protein of about 6 Kd. The beta
			subunit seems [2] to regulate, through the assembly of alpha/beta
			heterodimers, the number of sodium pumps transported to the
I	i .		plasma membrane.
		i	This family is typically found in association with E1 E2
			This family is typically found in association with E1-E2 ATPase. Uses of these polypeptide includes regulating that ion
			ATPase. Uses of these polypeptide includes regulating that ion
Na K ATPase N		Na+/K+ ATPase C-	ATPase. Uses of these polypeptide includes regulating that ion content in a desired cell or organism and can convey salt or ion



Pfam	Prosite	Full Name	049 Description
1 Itali		terminus	Definition: Na+/K+ ATPase C-terminus Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_138 (release 2.1) Gathering cutoffs: 15.6 15.6 Trusted cutoffs: 15.60 15.60 Noise cutoffs: 15.10 15.10 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference INTERPRO; IPR000661; Database reference: PFAMB; PB000031; Comment: PFAMB; PB000031; This family is always found in association with E1-E2_ATPase. Comment: This extension is specific to the Na+/K+ ATPase subfamily of Comment: ATPases. Number of members: 90
NAD_Gly3P_dh	PDOC00740	NAD-dependent glycerol- 3-phosphate dehydrogenase signature	NAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.1.8) (GPD) catalyzes the reversible reduction of dihydroxyacetone phosphate to glycerol-3-phosphate. It is a eukaryotic cytosolic homodimeric protein of about 40 Kd. As a signature pattern we selected a glycine-rich region that is probably [1] involved in NAD-binding. Description of pattern(s) and/or profile(s) Consensus pattern G-[AT]-[LIVM]-K-[DN]-[LIVM](2)-A-x-[GA]-x-G-[LIVMF]-x- [DE]-G-[LIVM]-x-{LIVMFYW}-G-x-N Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Pattern and text revised.
NifU_N		NifU-like N terminal domain	References [1] Otto J., Argos P., Rossmann M.G. Eur. J. Biochem. 109:325-330(1980). Accession number: PF01592 Definition: NifU-like N terminal domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_772 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 1.20 1.20 Noise cutoffs: -28.80 -28.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97032601 Reference Title: protein, is highly conserved in evolution. Reference Title: protein, is highly conserved in evolution. Hwang DM, Dempsey A, Tan KT, Liew CC; Reference Location: J Mol Evol 1996;43:536-540. INTERPRO; IPR002871; This domain is found in NifU in combination with NifU-like. Comment: This domain is found on isolated in several
			bacterial species Comment: such as Swiss:O53156. The nif genes are responsible for nitrogen Comment: fixation. However this domain is found in bacteria that do not Comment: fix nitrogen, so it may have a broader significance in the cell Comment: than nitrogen fixation.

Pfam	Prosite	Full Name	Description
			Number of members: 32
NLPC_P60		NLP/P60 family	Accession number: PF00877
			Definition: NLP/P60 family Author: Bateman A
			Alignment method of seed: HMM_built_from_alignment
			Source of seed members: Pfam-B 292 (release 3.0)
			Gathering cutoffs: -9 -9
			Trusted cutoffs: -8.30 -8.30
			Noise cutoffs: -10.40 -10.40
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Database Reference INTERPRO; IPR000064;
			Database reference: PFAMB; PB024706; Comment: PFAMB; PB024706; The function of this domain is unknown. It is
			found
			Comment: in several lipoproteins.
			Number of members: 54
NTR		NTR/C345C module	Accession number: PF01759
			Definition: NTR/C345C module
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: [1]
			Gathering cutoffs: 25 25
			Trusted cutoffs: 57.30 57.30
			Noise cutoffs: 2.80 2.80 HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 99379676
			Reference Title: The NTR module: domains of netrins,
			secreted frizzled
	-		Reference Title: related proteins, and type I procollagen C-
			proteinase
			Reference Title: enhancer protein are homologous with
			tissue inhibitors of
			Reference Title: metalloproteases [In Process Citation]
			Reference Author: Banyai L, Patthy L; Reference Location: Protein Sci 1999:8:1636-1642.
			Reference Location: Protein Sci 1999;8:1636-1642. Database Reference INTERPRO; IPR001134;
			Database reference: PFAMB; PB005955;
			Comment: We have not included the related TIMP
			family.
			Comment: It has been suggested that the common
			function of these
			Comment: modules is binding to metzincins [1]. A
			subset of this family
			Comment: is known as the C345C domain because it
			occurs in complement
			Comment: C3, C4 and C5.
			Number of members: 64
Nucleoside tran		Nucleoside transporter	Accession number: PF01733
			Definition: Nucleoside transporter
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_2135 (release 4.1)
			Gathering cutoffs: 25 25
			Trusted cutoffs: 25.50 25.50
			Noise cutoffs: -122.50 -122.50
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			Reference Number: [1] Reference Medline: 98148080
			Reference Title: Cloning of the human equilibrative,
			Reference Title: nitrobenzylmercaptopurine riboside
			(NBMPR)-insensitive
		1	Reference Title: nucleoside transporter ei by functional
			expression in a
		- 1	Reference Title: transport-deficient cell line.
			Reference Author: Crawford CR, Patel DH, Naeve C, Belt
	1		JA;
			Reference Location: J Biol Chem 1998;273:5288-5293.
	ĺ	1	Reference Number: [2]

			321
Pfam	Prosite	Full Name	Description Reference Medline: 98019212
			Reference Medline: 98019212 Reference Title: Molecular cloning and functional
			characterization of
			Reference Title: nitrobenzylthioinosine (NBMPR)-sensitive
			(es) and
			Reference Title: NBMPR-insensitive (ei) equilibrative
			nucleoside transporter
			Reference Title: proteins (rENT1 and rENT2) from rat tissues.
			Reference Author: Yao SY, Ng AM, Muzyka WR, Griffiths M,
			Cass CE, Baldwin SA,
			Reference Author: Young JD;
			Reference Location: J Biol Chem 1997;272:28423-28430.
			Database Reference INTERPRO; IPR002259;
			Comment: This is a family of nucleoside transporters. Comment: In mammalian cells nucleoside transporters
			Comment: In mammalian cells nucleoside transporters transport nucleoside
			Comment: across the plasma membrane and are
			essential for nucleotide
			Comment: synthesis via the salvage pathways for cells
			that lack their own
			Comment: de novo synthesis pathways [2].
			Comment: Also in this family is mouse and human
			nucleolar protein HNP36 Comment: Swiss:Q14542 a protein of unknown
			Comment: Swiss:Q14542 a protein of unknown function; although it has been
			Comment: hypothesized to be a plasma membrane
٠.			nucleoside transporter [2].
			Number of members: 15
0111100			
Orbi_VP6		Orbivirus helicase VP6	Accession number: PF01516
			Definition: Orbivirus helicase VP6 Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B 765 (release 4.0)
			Gathering cutoffs: -68 -68
			Trusted cutoffs: -37.10 -37.10
			Noise cutoffs: -98.90 -98.90
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1] Reference Medline: 97456481
			Reference Title: Bluetongue virus VP6 protein binds ATP
			and exhibits an
			Reference Title: RNA-dependent ATPase function and a
		:	helicase activity that
			Reference Title: catalyze the unwinding of double-stranded
			RNA substrates.
			Reference Author: Stauber N, Martinez-Costas J, Sutton G, Monastyrskaya K,
			Reference Author: Roy P;
			Reference Location: J Virol 1997;71:7220-7226.
			Database Reference INTERPRO; IPR001399;
			Comment: The VP6 protein a minor protein in the core
			of the virion
			Comment: is probably the viral helicase [1]. Number of members: 27
			radination of monipoles.
OSCP	PDOC00327	ATP synthase delta	ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2]
		(OSCP) subunit	is a component
		signature	of the cytoplasmic membrane of eubacteria, the inner membrane
			of mitochondria,
	!		and the thylakoid membrane of chloroplasts. The ATPase
			complex is composed of an oligomeric transmembrane sector, called CF(0), which acts
			as a proton
			channel, and a catalytic core, termed coupling factor CF(1).
			, , , , , , , , , , , , , , , , , , , ,
			One of the subunits of the ATPase complex, known as subunit
			delta in bacteria
			and chloroplasts or the Oligomycin Sensitivity Conferral Protein
			(OSCP) in
			mitochondria, seems to be part of the stalk that links CF(0) to CF(1). It
	l	L	Ot (1). It



			702
Pfam	Prosite	Full Name	Description either transmits conformational changes from CF(0) into CF(1) or is involved
			in proton conduction [3]. The different delta/OSCP subunits are proteins of approximately 200 amino-acid residues - once the transit peptide has been removed in the chloroplast and mitochondrial forms - which show only moderate sequence homology. The signature pattern used to detect ATPase delta/OSCP
			subunits is based on a conserved region in the C-terminal section of these proteins.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIVM]-x-[LIVMFYT]-x(3)-[LIVMT]-[DENQK]-x(2)-[LIVM]-x- [GSA]-G-[LIVMFYGA]-x-[LIVM]-[KRHENQ]-x-[GSEN] Sequences known to belong to this class detected by the pattern ALL, except 3 sequences. Other sequence(s) detected in SWISS-PROT 2. Last update November 1997 / Pattern and text revised. References [1] Futai M., Noumi T., Maeda M.
			Annu. Rev. Biochem. 58:111-136(1989). [2] Senior A.E. Physiol. Rev. 68:177-231(1988).
			[3] Engelbrecht S., Junge W. Biochim. Biophys. Acta 1015:379-390(1990).
OTCace	PDOC00091	Aspartate and ornithine carbamoyltransferases signature	Aspartate carbamoyltransferase (EC 2.1.3.2) (ATCase) catalyzes the conversion of aspartate and carbamoyl phosphate to carbamoylaspartate, the second step in the de novo biosynthesis of pyrimidine nucleotides [1]. In prokaryotes ATCase consists of two subunits: a catalytic chain (gene pyrB) and a regulatory chain (gene pyrI), while in eukaryotes it is a domain in a multifunctional enzyme (called URA2 in yeast, rudimentary in Drosophila, and CAD in mammals [2]) that also catalyzes other steps of the biosynthesis of pyrimidines.
			Ornithine carbamoyltransferase (EC 2.1.3.3) (OTCase) catalyzes the conversion of ornithine and carbamoyl phosphate to citrulline. In mammals this enzyme participates in the urea cycle [3] and is located in the mitochondrial matrix. In prokaryotes and eukaryotic microorganisms it is involved in the biosynthesis of arginine. In some bacterial species it is also involved in the degradation of arginine [4] (the arginine deaminase pathway).
			It has been shown [5] that these two enzymes are evolutionary related. The predicted secondary structure of both enzymes are similar and there are some regions of sequence similarities. One of these regions includes three



		5	953
Pfam	Prosite	Full Name	Description residues which have been shown, by crystallographic studies
			[6], to be implicated in binding the phosphoryl group of carbamoyl phosphate. We have selected this region as a signature for these enzymes.
			Description of pattern(s) and/or profile(s) Consensus pattern F-x-[EK]-x-S-[GT]-R-T [S, R, and the 2nd T
			bind carbamoyl phosphate] Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Note the residue in position 3 of the pattern allows to distinguish between an ATCase (Glu) and an OTCase (Lys). Last update October 1993 / Text revised. References [1] Lerner C.G., Switzer R.L. J. Biol. Chem. 261:11156-11165(1986).
			[2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).
			[3] Takiguchi M., Matsubasa T., Amaya Y., Mori M. BioEssays 10:163-166(1989).
			[4] Baur H., Stalon V., Falmagne P., Luethi E., Haas D. Eur. J. Biochem. 166:111-117(1987).
			[5] Houghton J.E., Bencini D.A., O'Donovan G.A., Wild J.R. Proc. Natl. Acad. Sci. U.S.A. 81:4864-4868(1981).
			[6] Ke HM., Honzatko R.B., Lipscomb W.N. Proc. Natl. Acad. Sci. U.S.A. 81:4037-4040(1984).
oxidored_q1_N		NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus	Accession number: PF00662 Definition: NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_22 (release 2.1) Gathering cutoffs: 18 18 Trusted cutoffs: 19.40 19.40 Noise cutoffs: 19.40 19.40 Noise cutoffs: 16.70 16.70 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Nedline: 93110040 The NADH:ubiquinone oxidoreductase (complex I) of Reference Title: respiratory chains. Walker JE; Reference Author: Walker JE; Reference Location: Database Reference: Database reference: PFAMB; PB000410; Database reference: Database reference: PFAMB; PB033295; Database reference: Database reference: Omment: ctrain 15 and Comment: eubacterial chain L are in this family.
			Comment: eubacterial chair Lare in this family. This sub-family is part of complex I which



	_	Ğ	954
Pfam	Prosite	Full Name	Description
			catalyses the Comment: transfer of two electrons from NADH to ubiquinone in a Comment: reaction that is associated with proton translocation Comment: across the membrane. Number of members: 546
oxidored_q2		NADH- ubiquinone/plastoquinon e oxidoreductase chain 4L	Accession number: PF00420 Definition: NADH-ubiquinone/plastoquinone oxidoreductase chain 4L Author: Finn RD Alignment method of seed: Clustalw Source of seed members: Pfam-B_193 (release 1.0) Gathering cutoffs: 25 15 Trusted cutoffs: 29.70 29.70 Noise cutoffs: 20.40 20.40 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference INTERPRO; IPR001133; Database reference: PFAMB; PB006066; Number of members: 219
PAN	PDOC00376	Apple domain	Plasma kallikrein (EC 3.4.21.34) and coagulation factor XI (EC 3.4.21.27) are two related plasma serine proteases activated by factor XIIA and which share the same domain topology: an N-terminal region that contains four tandem repeats of about 90 amino acids and a C-terminal catalytic domain. The 90 amino-acid repeated domain contains 6 conserved cysteines. It has been shown [1,2] that three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines. The domain can be drawn in the shape of an apple (see below) and has been accordingly called the 'apple domain'. XXX XXX X CC X X X X X X CXXX X X 1 X X X Schematic representation of an X CXXX X X Apple domain. X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X



			155
Pfam	Prosite	Full Name	Description
			[1] McMullen B.A., Fujikawa K., Davie E.W. Biochemistry 30:2050-2056(1991).
			[2] McMullen B.A., Fujikawa K., Davie E.W.
			Biochemistry 30:2056-2060(1991).
PAP2		PAP2 superfamily	Accession number: PF01569 Definition: PAP2 superfamily Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw Source of seed members: Pfam-B 486 (release 4.0)
			Gathering cutoffs: 16 16 Trusted cutoffs: 22.00 22.00
			Noise cutoffs: 11.40 11.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1] Reference Medline: 97194074
			Reference Title: Identification of a novel phosphatase sequence motif. Reference Author: Stukey J, Carman GM;
			Reference Location: Protein Sci 1997;6:469-472. Reference Number: [2]
			Reference Medline: 97406916 Reference Title: An unexpected structural relationship between integral
			Reference Title: membrane phosphatases and soluble haloperoxidases.
			Reference Author: Neuwald AF; Reference Location: Protein Sci 1997;6:1764-1767. Database Reference INTERPRO; IPR000326;
	:		Database reference: PFAMB; PB021113; Database reference: PFAMB; PB040926;
			Database reference: PFAMB; PB041096; Database reference: PFAMB; PB041301; Comment: PFAMB; PB041301; This family includes the enzyme type 2
			phosphatidic acid Comment: phosphatase (PAP2).
04.50		Dhasahaadaaaina	Number of members: 49 Accession number: PF01507
PAPS_reduct		Phosphoadenosine phosphosulfate reductase family	Accession number: PF01507 Definition: Phosphoadenosine phosphosulfate reductase family
			Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B 590 (release 4.0)
			Gathering cutoffs: 49 49 Trusted cutoffs: 55.40 55.40
			Noise cutoffs: -34.60 -34.60 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Medline: 97411695
			Reference Title: Crystal structure of phosphoadenylyl sulphate (PAPS) Reference Title: reductase: a new family of adenine
			nucleotide alpha Reference Title: hydrolases.
			Reference Author: Savage H, Montoya G, Svensson C, Schwenn JD, Sinning I; Reference Location: Structure 1997;5:895-906.
			Reference Number: [2] Reference Medline: 96061968
			Reference Title: Reaction mechanism of thioredoxin: 3'-phospho-adenylylsulfate reductase investigated by
			Reference Title: site-directed mutagenesis. Reference Author: Berendt U, Haverkamp T, Prior A,
			Schwenn JD; Reference Location: Eur J Biochem 1995;233:347-356. Reference Number: [3]
			Reference Medline: 91066949



D.C.	T	000000000000000000000000000000000000000	956
Pfam	Prosite	Full Name	Description
			Reference Title: ATP sulphurylase activity of the nodP and
			nodQ gene Reference Title: products of Rhizobium meliloti
			Reference Title: products of Rhizobium meliloti. Reference Author: Schwedock J, Long SR;
			Reference Location: Nature 1990;348:644-647.
			Database Reference: SCOP; 1sur; fa; [SCOP-USA][CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR002500;
			Database Reference PDB; 1sur ; 48; 215;
			Comment: This domain is found in phosphoadenosine
			phosphosulfate (PAPS) reductase Comment: enzymes or PAPS sulfotransferase PAPS
			Comment: enzymes or PAPS sulfotransferase. PAPS reductase is part of the adenine
			Comment: nucleotide alpha hydrolases superfamily
			also including N type ATP PPases
		•	Comment: and ATP sulphurylases [1]. The enzyme
			uses thioredoxin as an electron
			Comment: donor for the reduction of PAPS to phospho-
			adenosine-phosphate (PAP) [1,2]. Comment: It is also found in NodP nodulation protein P
			Comment: It is also found in NodP nodulation protein P from Rizobium which has ATP
			Comment: sulpurylase activity (sulfate adenylate
			transferase) [3].
			Number of members: 48
PARP		- · //	
PARP		Poly(ADP-ribose)	Accession number: PF00644
		polymerase catalytic region	Definition: Poly(ADP-ribose) polymerase catalytic region. Author: Bateman A
		region	Author: Bateman A Alignment method of seed: HMM_built_from_alignment
			Source of seed members: Bateman A
İ			Gathering cutoffs: -59.4 -59.4
			Trusted cutoffs: -44.60 -44.60
İ			Noise cutoffs: -180.60 -180.60
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 96353841
	į		Reference Title: Structure of the catalytic fragment of poly(AD-ribose)
			Reference Title: polymerase from chicken.
	İ		Reference Author: Ruf A, Mennissier de Murcia J, de Murcia
			G, Schulz GE;
			Reference Location: Proc Natl Acad Sci U S A 1996;93:7481-
	-		7485.
			Reference Number: [2]
			Reference Medline: 93293867 Reference Title: The carboxyl-terminal domain of human
i			Reference Title: The carboxyl-terminal domain of human poly(ADP-ribose)
	1		Reference Title: polymerase. Overproduction in Escherichia
			coli, large scale
	1		Reference Title: purification, and characterization.
			Reference Author: Simonin F, Hofferer L, Panzeter PL,
			Muller S, de Murcia G,
			Reference Author: Althaus FR;
			Reference Location: J Biol Chem 1993;268:13454-13461. Database Reference: SCOP: 1paw: fa: ISCOP-USAUCATH.
			Database Reference: SCOP; 1paw; fa; [SCOP-USA][CATH-PDBSUM]
			Database Reference INTERPRO; IPR001290;
			Database Reference PDB; 1a26; 662; 997;
			Database Reference PDB; 1pax; 662; 997;
			Database Reference PDB; 2pax; 662; 997;
			Database Reference PDB; 3pax; 662; 997;
			Database Reference PDB; 4pax ; 662; 997;
			Database Reference PDB; 2paw; 662; 1009;
			Database reference: PFAMB; PB041409; Comment: Poly(ADP-ribose) polymerase catalyses the
			Comment: Poly(ADP-ribose) polymerase catalyses the covalent
			Comment: attachment of ADP-ribose units from NAD+
			to itself and
			Comment: to a limited number of other DNA binding
	İ		proteins, which
			Comment: decreases their affinity for DNA.
			Comment: Poly(ADP-ribose) polymerase is a
			regulatory component



	<u> </u>	957	
Pfam Prosite	Full Name	Description	
		l .	induced by DNA damage.
		Comment:	The carboxyl-terminal region is the most
		highly conserved	to of the markets. Experience to be use
		Comment:	region of the protein. Experiments have
		shown that a	carboxyl 40 kDa fragment is still catalytically
		active [2].	Salbonyi to Rea naginoni io siii oataiyioany
		Number of members:	19
PC_rep	Proteasome/cyclosome	Accession number:	PF01851
	repeat		roteasome/cyclosome repeat
			ateman A
		Alignment method of Source of seed meml	
		Gathering cutoffs:	25 0
		Trusted cutoffs:	30.60 3.00
			15.80 15.80
			line: hmmbuild HMM SEED
		HMM build command	line: hmmcalibrateseed 0 HMM
		Reference Number:	[1]
		Reference Medline:	97348748
		Reference Title:	A repetitive sequence in subunits of the
	1	26S proteasome and	20S cyclosome (anaphase-promoting
		Reference Title:	203 cyclosome (anaphase-promoting
		complex). Reference Author:	Lupas A, Baumeister W, Hofmann K;
		Reference Location:	Trends Biochem Sci 1997;22:195-196.
		Database Reference	INTERPRO; IPR002015;
		Database reference:	PFAMB; PB009978;
		Database reference:	PFAMB; PB040656;
		Number of members:	112
	DE ("	A	DE00024
PE	PE family	Accession number:	PF00934
			E family ateman A
		Alignment method of	
			bers: Pfam-B 253 (release 3.0)
		Gathering cutoffs:	-20 -20
		Trusted cutoffs:	-10.80 -10.80
		Noise cutoffs:	-20.60 -20.60
			line: hmmbuild HMM SEED
			l line: hmmcalibrateseed 0 HMM
		Reference Number:	[1] 98295987
		Reference Medline: Reference Title:	Deciphering the biology of Mycobacterium
		tuberculosis from	Seephering the blology of Mycobacterium
		Reference Title:	the complete genome sequence.
		Reference Author:	Cole ST, Brosch R, Parkhill J, Garnier T,
		Churcher C,	
		Reference Author:	Harris D, Gordon SV, Eiglmeier K, Gas S,
		Barry CE 3rd, Reference Author:	Tekaia F, Badcock K, Basham D, Brown
		D, Chillingworth T,	Tonaia I , Dadoook II, Daonain D, Diowii
1		Reference Author:	Connor R, Davies R, Devlin K, Feltwell T,
		Gentles S, Hamlin	
		Reference Author:	N, Holroyd S, Hornsby T, Jagels K, Barrell
		BG, et al;	N
		Reference Location:	Nature 1998;393:537-544
		Database Reference	INTERPRO; IPR000084; This family named after a PE motif near to
		Comment: the amino	This family hamed after a FE mote flear to
		Comment:	terminus of the domain. The PE family of
		proteins	torring of the derivative file of a family of
		Comment:	all contain an amino-terminal region of
		about 110	_
		Comment:	amino acids. The carboxyl terminus of this
		family	and the land of the land of the land
		Comment:	are variable and fall into several classes.
			largest class of PE proteins is the highly
		Comment:	largest class of PE proteins is the highly
			largest class of PE proteins is the highly PGRS class which have a high glycine
		Comment: repetitive	

			958
Pfam	Prosite .	Full Name	Description
			but it
			Comment: has been suggested that they may be
			related to
			Comment: antigenic variation of Mycobacterium
			tuberculosis [1]. Number of members: 90
Pep deformylase		Polypeptide deformylase	Accession number: PF01327
'- '		'' '	Definition: Polypeptide deformylase
		1	Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Sarah Teichmann
			Gathering cutoffs: 25 25
			Trusted cutoffs: 157.40 157.40 Noise cutoffs: -29.00 -29.00
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 97002011
			Reference Title: A new subclass of the zinc
			metalloproteases superfamily
			Reference Title: revealed by the solution structure of peptide
			deformylase. Reference Author: Mainpel T. Blanquet S. Dardel F:
			Reference Author: Meinnel T, Blanquet S, Dardel F; Reference Location: J Mol Biol 1996;262:375-386.
			Reference Location: J Moi Biol 1996;262.375-366. Reference Number: [2]
			Reference Medline: 98332750
			Reference Title: Solution structure of nickel-peptide
			deformylase.
			Reference Author: Dardel F, Ragusa S, Lazennec C,
		İ	Blanquet S, Meinnel T;
			Reference Location: J Mol Biol 1998;280:501-513.
			Database Reference: SCOP; 1def; fa; [SCOP-USA][CATH-
			PDBSUM] Database Reference INTERPRO; IPR000181;
			Database Reference PDB; 2def; 4; 142;
			Database Reference PDB; 1def; 4; 142;
			Database Reference PDB; 1dff; 4; 142;
			Database Reference PDB; 1bsj A; 4; 142;
	1		Database Reference PDB; 1bsk A; 4; 142;
	1		Database Reference PDB; 1bs4 A; 4; 142;
			Database Reference PDB; 1bs4 B; 504; 642; PDB; 1bs4 C; 1004; 1142;
			Database Reference PDB; 1bs5 A; 4; 142;
			Database Reference PDB; 1bs5 B; 504; 642;
			Database Reference PDB; 1bs5 C; 1004; 1142;
			Database Reference PDB; 1bs6 A; 4; 142;
			Database Reference PDB; 1bs6 B; 504; 642;
ı			Database Reference PDB; 1bs6 C; 1004; 1142;
ı			Database Reference PDB; 1bs7 A; 4; 142;
			Database Reference PDB; 1bs7 B; 504; 642;
I			Database Reference PDB; 1bs7 C; 1004; 1142; PDB; 1bs8 A; 4; 142;
	1		Database Reference PDB; 1bs8 B; 504; 642;
	1		Database Reference PDB; 1bs8 C; 1004; 1142;
			Database Reference PDB; 1bsz A; 4; 142;
			Database Reference PDB; 1bsz B; 504; 642;
			Database Reference PDB; 1bsz C; 1004; 1142;
			Database Reference PDB; 1icj A; 4; 142;
			Database Reference PDB; 1icj B; 504; 642;
			Database Reference PDB; 1icj C; 1004; 1142; PFAMB; PB041251;
			Number of members: 25
Peptidase C15		Pyroglutamyl peptidase	Accession number: PF01470
. 55114400_010		, ,	Definition: Pyroglutamyl peptidase
			Author: Bateman A
			Alignment method of seed: Clustalw_manual
			Source of seed members: [1]
			Gathering cutoffs: 25 25
1			Trusted cutoffs: 436.10 436.10 Noise cutoffs: -155.40 -155.40
			Noise cutoffs: -155.40 -155.40 HMM build command line: hmmbuild HMM SEED
	1		HMM build command line: hmmcalibrateseed 0 HMM



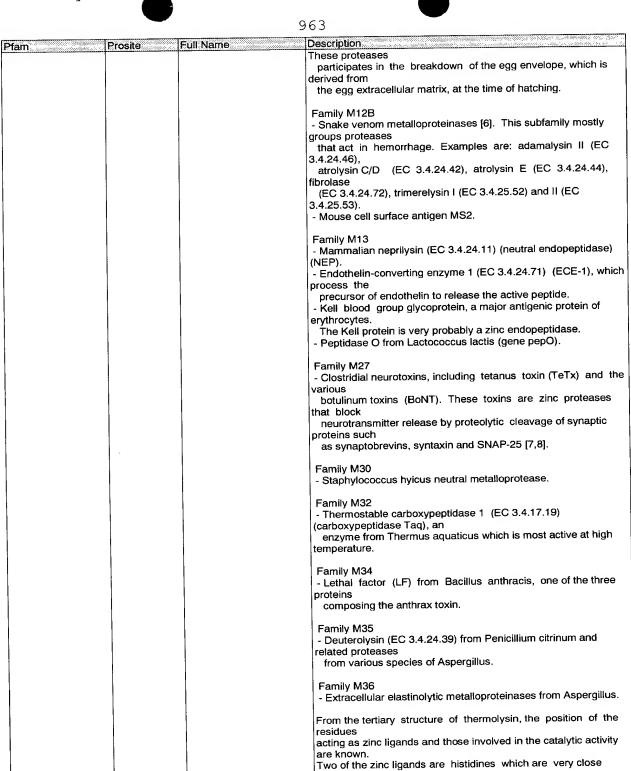
Pfam				
Pfam	Prosite	Full Name	Description	March 1882 Bloom and the company of the contract of the contra
			Reference Number:	[1]
			Reference Medline:	99216536
			Reference Title: peptidase I from	The crystal structure of pyroglutamyl
			Reference Title:	hacillus amulalisustasians anna la
		1	structure for a	bacillus amyloliquefaciens reveals a new
			Reference Title:	cysteine protease.
		1	Reference Author:	Odagaki Y, Hayashi A, Okada K, Hirotsu
			K, Kabashima T, Ito	Oddgam 1, Hayashi A, Okada N, Hiloisu
			Reference Author:	K, Yoshimoto T, Tsuru D, Sato M, Clardy
			J	, reality, round B, odio M, oldrdy
			Reference Location:	Structure 1999;7:399-411.
			Database Reference:	SCOP; 1aug; fa; [SCOP-USA][CATH-
			PDBSUM]	
			Database Reference	MEROPS; C15;
ł			Database Reference	INTERPRO; IPR000816;
			Database Reference Database Reference	PDB; 1a2z A; 2; 209;
		1	Database Reference	PDB; 1a2z B; 2; 209;
			Database Reference	PDB; 1a2z C; 2; 209; PDB; 1a2z D; 2; 209;
		1	Database Reference	PDB; 1aug A; 3; 204;
	1		Database Reference	PDB; 1aug B; 213; 414;
	1		Database Reference	PDB; 1aug C; 423; 624;
			Database Reference	PDB; 1aug D; 633; 834;
			Number of members:	10
Peptidase_M20	PDOC00613	AraE / danE / AOV/4 /	 	
· optiduso_ivizo	1100000013	ArgE / dapE / ACY1 / CPG2 / yscS family	The following enzyme	es have been shown [1,2,3] to be
		signatures	evolutionary and Functionally related:	
		orginata Co	Functionally related:	
			- In the biosynthetic p	athway from glutamate to arginine, the
			removal of an	atiway non gidiamate to arginine, the
				2-acetylornithine can be catalyzed via two
			distinct	2 assetyler million carr be catalyzed via two
			enzymatic strategies	s depending on the organism. In some
			bacteria and in	
			fungi, the acetyl gr	roup is transferred on glutamate by
			glutamate	_
			acetyltransferase (EC	C 2.3.1.35) while in enterobacteria such as
			Escherichia	bu part de militare de la companya d
			3.5.1.16)	by acetylornithine deacetylase (EC
				AO) (gene argE). AO is a homodimeric
			cobalt-dependent	(a) (gone argu). No is a nomodifient
			enzyme which displa	ys broad specificity and can also
			deacylates substrates	
			such as acetylarginin	e, acetylhistidine, acetylglutamate
			semialdehyde, etc.	
			- Succinyidiaminopime	elate desuccinylase (EC 3.5.1.18) (SDAP)
			(gene dapE) is	atalyzas the fifth stee in the bissessite i
			of lysine	atalyzes the fifth step in the biosynthesis
		•		aldehyde: the hydrolysis of succinyl-
			diaminopimelate to	
			diaminopimelate and	succinate. SDAP is an enzyme that
			requires cobalt or	•
	j		zinc as a cofactor.	
			- Aminoacylase-1 [4]	(EC 3.5.1.14) (N-acyl-l-amino-acid
	ļ		amidohydrolase)	amadia di di di di
	ļ		enzyme that catalyzes	omodimeric zinc-binding mammalian
				pha-acylated amino acids (except for
			aspartate).	prix adjusted arrillo acids (except for
				2 (EC 3.4.17.11) (folate hydrolase G2)
			(gene cpg2) from	
			Pseudomonas strain	RS-16. This enzyme catalyzes the
			hydrolysis of reduced	
			and non-reduced fola	ites to pteroates and glutamate. G2 is a
			homodimeric	
1			zinc-dependent enzyr	ne.
			- vacuolar carboxypept (gene CPS1).	tidase S (EC 3.4.17.4) (yscS) from yeast
				1.11) (gene pepT) (tripeptidase) from
			bacteria. This	/ (gene pep i) (inhehilidase) itom



			960
Pfam	Prosite	Full Name	Description enzyme catalyzes a variety of tripeptides containing N-terminal methionine, leucine, or phenylalanine Xaa-His dipeptidase (EC 3.4.13.3) (carnosinase) from Lactobacillus (gene pepV) [5], a metalloenzyme with activity against beta-alanyl-dipeptides including carnosine (beta-alanyl-histidine). These enzymes share a few characteristics. They hydrolyse peptidic bonds in Substrates that share a common structure, they are dependent on cobalt or zinc For their activity and they are proteins of 40 Kd to 60 Kd with a number of Regions of sequence similarity. As signature patterns for these proteins, we selected two of the conserved Regions. The first pattern contains a conserved histidine which could be Involved in binding metal ions and the second pattern contains a number of Conserved charged residues.
			Description of pattern(s) and/or profile(s) Consensus pattern [LIV]-[GALMY]-[LIVMF]-x-[GSA]-H-x-D-[TV]-[STAV] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 6. Consensus pattern [GSTAI]-[SANQ]-D-x-K-[GSACN]-x(2)-[LIVMA]-x(2)-[LIVMFY]- x(14,17)-[LIVM]-x-[LIVMF]-[LIVMSTAG]-[LIVMFA]-x(2)-[DNG]- E-E-x-[GSTN] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note these proteins belong to families M20A/M20B in the classification of peptidases [6,E1]. Last update November 1997 / Patterns and text revised. References [1] Meinnel T., Schmitt E., Mechulam Y., Blanquet S. J. Bacteriol. 174:2323-2331(1992).
			[2] Boyen A., Charlier D., Sakanyan V., Mett I., Glansdorff N. Gene 116:1-6(1992). [3] Miller C.G., Miller J.L., Bagga D.A. J. Bacteriol. 173:3554-3558(1991). [4] Mitta M., Ohnogi H., Yamamoto A., Kato I., Sakiyama F., Tsunasawa S. J. Biochem. 112:737-742(1992). [5] Vongerichten K., Klein J., Matern H., Plapp R.
			Microbiology 140:2591-2600(1994). [6] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995). [E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt

			961
Pfam	Prosite	Full Name	Description
Peptidase_M3	PDOC00129	Neutral zinc metallopeptidases, zinc- binding region signature	The majority of zinc-dependent metallopeptidases (with the notable exception Of the carboxypeptidases) share a common pattern of primary structure [1,2,3] in the part of their sequence involved in the binding of zinc, and can be grouped together as a superfamily,known as the metzincins, or the basis of this sequence similarity. They can be classified into a number of distinct families [4,E1] which are listed below along with the proteases which are
			currently known to belong to these families. Family M1 - Bacterial aminopeptidase N (EC 3.4.11.2) (gene pepN). - Mammalian aminopeptidase N (EC 3.4.11.2). - Mammalian glutamyl aminopeptidase (EC 3.4.11.7) (aminopeptidase A). It may play a role in regulating growth and differentiation of early B-lineage cells. - Yeast aminopeptidase yscII (gene APE2). - Yeast alanine/arginine aminopeptidase (gene AAP1). - Yeast hypothetical protein YIL137c.
			- Leukotriene A-4 hydrolase (EC 3.3.2.6). This enzyme is responsible for the hydrolysis of an epoxide moiety of LTA-4 to form LTB-4; it has been shown that it binds zinc and is capable of peptidase activity. Family M2 - Angiotensin-converting enzyme (EC 3.4.15.1) (dipeptidyl carboxypeptidase I) (ACE) the enzyme responsible for hydrolyzing angiotensin I to angiotensin II. There are two forms of ACE: a testis-specific isozyme and
			a somatic isozyme which has two active centers. Family M3 - Thimet oligopeptidase (EC 3.4.24.15), a mammalian enzyme involved in the cytoplasmic degradation of small peptides Neurolysin (EC 3.4.24.16) (also known as mitochondrial oligopeptidase M or microsomal endopeptidase) Mitochondrial intermediate peptidase precursor (EC 3.4.24.59) (MIP). It is involved the second stage of processing of some proteins imported in the mitochondrion.
			- Yeast saccharolysin (EC 3.4.24.37) (proteinase yscD) Escherichia coli and related bacteria dipeptidyl carboxypeptidase (EC 3.4.15.5) (gene dcp) Escherichia coli and related bacteria oligopeptidase A (EC 3.4.24.70) (gene opdA or prlC) Yeast hypothetical protein YKL134c. Family M4 - Thermostable thermolysins (EC 3.4.24.27), and related
			thermolabile neutral proteases (bacillolysins) (EC 3.4.24.28) from various species of Bacillus. - Pseudolysin (EC 3.4.24.26) from Pseudomonas aeruginosa (gene lasB). - Extracellular elastase from Staphylococcus epidermidis. - Extracellular protease prt1 from Erwinia carotovora. - Extracellular minor protease smp from Serratia marcescens. - Vibriolysin (EC 3.4.24.25) from various species of Vibrio. - Protease prtA from Listeria monocytogenes.

	9	062
Pfam Prosite	Full Name	Description
		- Extracellular proteinase proA from Legionella pneumophila.
		Comilly ME
		Family M5 - Mycolysin (EC 3.4.24.31) from Streptomyces cacaoi.
		Myddydin (Ed d. 1.2 1.61) ffdin difoptomyddd diddiol
		Family M6
		- Immune inhibitor A from Bacillus thuringiensis (gene ina). Ina
		degrades two
		classes of insect antibacterial proteins, attacins and cecropins.
		Family M7
		- Streptomyces extracellular small neutral proteases
		Family M8
		- Leishmanolysin (EC 3.4.24.36) (surface glycoprotein gp63), a cell surface
		protease from various species of Leishmania.
		,
		Family M9
		- Microbial collagenase (EC 3.4.24.3) from Clostridium
	,	perfringens and Vibrio alginolyticus.
		alginoryticus.
		Family M10A
		- Serralysin (EC 3.4.24.40), an extracellular metalloprotease from
		Serratia.
		- Alkaline metalloproteinase from Pseudomonas aeruginosa (gene aprA).
	1	- Secreted proteases A, B, C and G from Erwinia chrysanthemi.
		- Yeast hypothetical protein YIL108w.
		_ "
		Family M10B
		- Mammalian extracellular matrix metalloproteinases (known as matrixins) [5]:
		MMP-1 (EC 3.4.24.7) (interstitial collagenase), MMP-2 (EC
		3.4.24.24) (72 Kd
		gelatinase), MMP-9 (EC 3.4.24.35) (92 Kd gelatinase), MMP-7
		(EC 3.4.24.23)
		(matrylisin), MMP-8 (EC 3.4.24.34) (neutrophil collagenase), MMP-3
		(EC 3.4.24.17) (stromelysin-1), MMP-10 (EC 3.4.24.22)
		(stromelysin-2), and
		MMP-11 (stromelysin-3), MMP-12 (EC 3.4.24.65) (macrophage
		metalloelastase).
		- Sea urchin hatching enzyme (envelysin) (EC 3.4.24.12). A protease that
		allows the embryo to digest the protective envelope derived
		from the egg
		extracellular matrix.
		- Soybean metalloendoproteinase 1.
		Family M11
		- Chlamydomonas reinhardtii gamete lytic enzyme (GLE).
		- "
		Family M12A
		- Astacin (EC 3.4.24.21), a crayfish endoprotease Meprin A (EC 3.4.24.18), a mammalian kidney and intestinal
		brush border
		metalloendopeptidase.
		- Bone morphogenic protein 1 (BMP-1), a protein which induces
		cartilage and
		bone formation and which expresses metalloendopeptidase activity. The
		Drosophila homolog of BMP-1 is the dorsal-ventral
		patterning protein
		tolloid.
		- Blastula protease 10 (BP10) from Paracentrotus lividus and
		the related protein SpAN from Strongylocentrotus purpuratus.
		- Caenorhabditis elegans protein toh-2.
		- Caenorhabditis elegans hypothetical protein F42A10.8.
		- Choriolysins L and H (EC 3.4.24.67) (also known as
		embryonic hatching
		proteins LCE and HCE) from the fish Oryzias lapides.



Two of the zinc ligands are histidines which are very close

together in the sequence; C-terminal to the first histidine is a glutamic acid residue which

acts as a nucleophile and promotes the attack of a water molecule on the

carbonyl carbon of the substrate. A signature pattern which includes the two

histidine and the glutamic acid residues is sufficient to detect

superfamily of proteins.



loc local		964
Pfam Pro	site Full Name	Description
		Description of pattern(s) and/or profile(s)
		Consensus pattern [GSTALIVN]-x(2)-H-E-[LIVMFYW]-{DEHRKP}-H-x-[LIVMFYWGSPQ] [The two H's are zinc ligands] [E is the active site residue]
		Sequences known to belong to this class detected by the pattern ALL, except for members of families M5, M7 amd M11. Other sequence(s) detected in SWISS-PROT 57; including Neurospora crassa conidiation-specific protein 13 which could be a zinc-protease. Last update
		July 1999 / Text revised. References [1] Jongeneel C.V., Bouvier J., Bairoch A. FEBS Lett. 242:211-214(1989).
		[2] Murphy G.J.P., Murphy G., Reynolds J.J. FEBS Lett. 289:4-7(1991).
		[3] Bode W., Grams F., Reinemer P., Gomis-Rueth FX., Baumann U., McKay D.B., Stoecker W. Zoology 99:237-246(1996).
	·	[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).
		[5] Woessner J. Jr. FASEB J. 5:2145-2154(1991).
		[6] Hite L.A., Fox J.W., Bjarnason J.B. Biol. Chem. Hoppe-Seyler 373:381-385(1992).
		[7] Montecucco C., Schiavo G. Trends Biochem. Sci. 18:324-327(1993).
		[8] Niemann H., Blasi J., Jahn R. Trends Cell Biol. 4:179-185(1994).
		[E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt
Peptidase_M48	Peptidase family M48	Accession number: PF01435 Definition: Peptidase family M48 Author: Bateman A
		Alignment method of seed: Clustalw_manual Source of seed members: Swiss-Prot Gathering cutoffs: -35 -35 Trusted cutoffs: -34.00 -34.00
		Noise cutoffs: -42.20 -42.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
		Database Reference MEROPS; M48; Database Reference: INTERPRO; IPR001915; Database reference: PFAMB; PB008839; Database reference: PFAMB; PB041497; Number of members: 28
Peptidase_S24	Peptidase family S24	Accession number: PF00717 Definition: Peptidase family S24 Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_616 (release 2.1) Gathering cutoffs: 1 1
		Gathering cutoffs: 1 1 Trusted cutoffs: 2.00 2.00 Noise cutoffs: -9.00 -9.00



		-	965
Pfam	Prosite	Full Name	Description HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference Database Reference: SCOP; 1umu; fa; [SCOP-USA][CATH-PDBSUM] Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference Database Reference PDB; 1umu B; 32; 123; PDB; 1ay9 B; 32; 123; PFAMB; PB05958; Database reference: PFAMB; PB041146; PFAMB; PB041823; 42
Peptidase_S8	PDOC00125	Serine proteases, subtilase family, active sites	Subtiliases [1,2] are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases and can be used as signatures specific to that category of proteases. The subtiliase family currently includes the following proteases: - Subtilisins (EC 3.4.21.62), these alkaline proteases from various Bacillus species have been the target of numerous studies in the past thirty years. - Alkaline elastase YaB from Bacillus sp. (gene ale) Alkaline serine exoprotease A from Vibrio alginolyticus (gene proA) Aqualysin I from Thermus aquaticus (gene pstl) AspA from Aeromonas salmonicida Bacillopeptidase F (esterase) from Bacillus subtilis (gene bpf) C5A peptidase F (esterase) from Bacillus subtilis (gene bpf) Cell envelope-located proteases PI, PII, and PIII from Lactococcus lactis Extracellular serine protease from Serratia marcescens Extracellular serine protease from Serratia marcescens Extracellular protease from Xanthomonas campestris Intracellular serine protease vpr from Bacillus subtilis (gene vpr) Minor extracellular serine protease vpr from Bacillus subtilis (gene vpr) Nisin leader peptide processing protease nisP from Lactococcus lactis Serotype-specific antigene 1 from Pasteurella haemolytica (gene spa1) Thermitase (EC 3.4.21.66) from Thermoactinomyces vulgaris Calcium-dependent protease from Anabaena variabilis (gene prcA) Halolysin from halophilic bacteria sp. 172p1 (gene hly) Alkaline extracellular protease (AEP) from Yarrowia lipolytica (gene PRB1) Cerevisin (EC 3.4.21.48) (vacuolar protease B) from yeast (gene PRB1).

			966
Pfam	Prosite	Full Name	Description
·			 - Kexin (EC 3.4.21.61) from yeast (gene KEX-2). - Oryzin (EC 3.4.21.63) (alkaline proteinase) from Aspergillus (gene alp). - Proteinase K (EC 3.4.21.64) from Tritirachium album (gene proK).
			- Proteinase R from Tritirachium album (gene proR) Proteinase T from Tritirachium album (gene proT) Subtilisin-like protease III from yeast (gene YSP3) Thermomycolin (EC 3.4.21.65) from Malbranchea sulfurea.
			- Furin (EC 3.4.21.85), neuroendocrine convertases 1 to 3 (NEC-1 to -3) and
			PACE4 protease from mammals, other vertebrates, and invertebrates. These
			proteases are involved in the processing of hormone precursors at sites comprised of pairs of basic amino acid residues [3].
			- Tripeptidyl-peptidase II (EC 3.4.14.10) (tripeptidyl aminopeptidase) from Human.
			- Prestalk-specific proteins tagB and tagC from slime mold [4]. Both proteins consist of two domains: a N-terminal subtilase catalytic
			domain and a C- terminal ABC transporter domain (see <pdoc00185>).</pdoc00185>
			Description of pattern(s) and/or profile(s)
			Consensus pattern [STAIV]-x-[LIVMF]-[LIVM]-D-[DSTA]-G- [LIVMFC]-x(2,3)-[DNH] [D is the active site residue] Sequences known to belong to this class detected by the pattern the majority of subtilases with a few exceptions. Other sequence(s) detected in SWISS-PROT 44.
			Consensus pattern H-G-[STM]-x-[VIC]-[STAGC]-[GS]-x-[LIVMA]- [STAGCLV]-[SAGM] [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for aspA and ssa1 which both seem to lack the histidine active site. Other sequence(s) detected in SWISS-PROT adenylate cyclase type VIII.
			Consensus pattern G-T-S-x-[SA]-x-P-x(2)-[STAVC]-[AG] [S is the active site residue] Sequences known to belong to this class detected by the pattern
			ALL, except for nisP, tagC and S.marcescens extracellular serine protease. Other sequence(s) detected in SWISS-PROT 6.
		ŀ	Note if a protein includes at least two of the three active site signatures, the probability of it being a serine protease from the subtilase family is 100%
			Note these proteins belong to family S8 in the classification of peptidases [5,E1]. Expert(s) to contact by email Brannigan J. jab5@vaxa.york.ac.uk
			Siezen R.J. siezen@nizo.n/
			Last update November 1997 / Patterns and text revised. References [1] Siezen R.J., de Vos W.M., Leunissen J.A.M., Dijkstra B.W.
			Protein Eng. 4:719-737(1991).
			Siezen R.J. In) Proceeding subtilisin symposium, Hamburg, (1992).
			3]

			967
Pfam	Prosite	Full Name	Description
			Barr P.J. Cell 66:1-3(1991).
			Cell 66:1-3(1991).
			[4]
			Shaulsky G., Kuspa A., Loomis W.F.; Genes Dev. 9:1111-1122(1995).
			defies bev. 9.1111-1122(1995).
			[5]
			Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).
			11001. E11291101. 244.15-61(1994).
			[E1]
			http://www.expasy.ch/cgi-bin/lists?peptidas.txt
Peptidase_S9	PDOC00587	Prolyl oligopeptidase	The prolyl oligopeptidase family [1,2,3] consist of a number of
		family serine active site	evolutionary
			related peptidases whose catalytic activity seems to be provided by a charge
			relay system similar to that of the trypsin family of serine
			proteases, but
			which evolved by independent convergent evolution. The known members of this
			family are listed below.
			- Prolyl endonentidage (EC 2 4 04 00) (DE)
			- Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post- proline cleaving
			enzyme). PE is an enzyme that cleaves peptide bonds on the
			C-terminal side
			of prolyl residues. The sequence of PE has been obtained from a mammalian
			species (pig) and from bacteria (Flavobacterium
			meningosepticum and Aeromonas hydrophila); there is a high degree of sequence
			conservation
			between these sequences.
			- Escherichia coli protease II (EC 3.4.21.83) (oligopeptidase B) (gene prtB)
			which cleaves peptide bonds on the C-terminal side of lysyl
			and argininyl residues.
			- Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an
			enzyme that
			removes N-terminal dipeptides sequentially from polypeptides having
			unsubstituted N-termini provided that the penultimate residue is
			proline.
			- Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: STE13) which is
			responsible for the proteolytic maturation of the alpha-factor
			precursor.
			- Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: DAP2).
			- Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide
j	1		hydrolase). This enzyme catalyzes the hydrolysis of the amino-terminal
	[peptide bond of
			an N-acetylated protein to generate a N-acetylated amino acid
	İ		and a protein with a free amino-terminus.
			A conserved serine residue has experimentally been shown (in E.coli protease
			l as well as in pig and bacterial PE) to be necessary for the
	1	! (catalytic
			mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp),
		į	s generally located about 150 residues away from the C-terminal
		6	extremity of
		l _t	hese enzymes (which are all proteins that contains about 700 o 800 amino
	[acids).
		-	



		9	968
Pfam	Prosite	Full Name	Description Description of pattern(s) and/or profile(s)
			Consensus pattern D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for yeast DPAP A. Other sequence(s) detected in SWISS-PROT NONE.
,			Note these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4,E1]. Last update November 1997 / Text revised.
			References [1]
			Rawlings N.D., Polgar L., Barrett A.J. Biochem. J. 279:907-911(1991).
			[2] Barrett A.J., Rawlings N.D. Biol. Chem. Hoppe-Seyler 373:353-360(1992).
			[3] Polgar L., Szabo E. Biol. Chem. Hoppe-Seyler 373:361-366(1992).
			[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).
			[E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt
Peptidase_U7		Peptidase family U7	Accession number: PF01343 Definition: Peptidase family U7 Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_707 (release 2.1) Gathering cutoffs: 25 25 Trusted cutoffs: 47.60 47.60 Noise cutoffs: -55.60 -55.60 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference MEROPS; U7; Database Reference INTERPRO; IPR002142; Number of members: 37
PEP-utilizers	PDOC00527	PEP-utilizing enzymes signatures	A number of enzymes that catalyze the transfer of a phosphoryl group from phosphoenolpyruvate (PEP) via a phospho-histidine intermediate have been shown to be structurally related [1,2,3,4]. These enzymes are:
			- Pyruvate, orthophosphate dikinase (EC 2.7.9.1) (PPDK). PPDK catalyzes the reversible phosphorylation of pyruvate and phosphate by ATP to PEP and diphosphate. In plants PPDK function in the direction of the
			formation of PEP, which is the primary acceptor of carbon dioxide in C4 and crassulacean
			acid metabolism plants. In some bacteria, such as Bacteroides symbiosus, PPDK functions in the direction of ATP synthesis Phosphoenolpyruvate synthase (EC 2.7.9.2) (pyruvate,water
			dikinase). This enzyme catalyzes the reversible phosphorylation of pyruvate by ATP to form
			PEP, AMP and phosphate, an essential step in gluconeogenesis when pyruvate and lactate are used as a carbon source Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9).
			This is the first enzyme of the phosphoenolpyruvate-dependent sugar phosphotransferase



Pfam	Prosite	Full Name	Description system (PTS), a major carbohydrate transport system in
			bacteria. The PTS catalyzes the phosphorylation of incoming sugar substrates
			concomitant with their translocation across the cell membrane. The general
			mechanism
			of the PTS is the following: a phosphoryl group from PEP is transferred
			to enzyme-I (EI) of PTS which in turn transfers it to a phosphoryl carrier
			protein (HPr). Phospho-HPr then transfers the phosphoryl group to a sugar-
			specific permease.
			All these enzymes share the same catalytic mechanism: they bind PEP and
			transfer the phosphoryl group from it to a histidine residue. The
			sequence around that residue is highly conserved and can be used as a
			signature pattern for these enzymes. As a second signature pattern we selected
			a conserved region in the C-terminal part of the PEP-utilizing enzymes. The
			biological
			significance of this region is not yet known.
			Description of pattern(s) and/or profile(s)
			Consensus pattern G-[GA]-x-[STN]-x-H-[STA]-[STAV]-[LIVM](2)-
			[STAV]-[RG] [H is phosphorylated] Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern [DEQSK]-x-[LIVMF]-S-[LIVMF]-G-[ST]-N-D- [LIVM]-x-Q- [LIVMFYGT]-[STALIV]-[LIVMFY]-[GAS]-x(2)-R Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update December 1999 / Patterns and text revised. References
			[1] Reizer J., Hoischen C., Reizer A., Pham T.N., Saier M.H. Jr. Protein Sci. 2:506-521(1993).
			[2] Reizer J., Reizer A., Merrick M.J., Plunkett G. III, Rose D.J., Saier M.H. Jr. Gene 181:103-108(1996).
			[3] Pocalyko D.J., Carroll L.J., Martin B.M., Babbitt P.C., Dunaway-Mariano D.
			Biochemistry 29:10757-10765(1990).
			[4] Niersbach M., Kreuzaler F., Geerse R.H., Postma P., Hirsch H.J. Mol. Gen. Genet. 232:332-336(1992).
PG_binding_2		Putative peptidoglycan binding domain	Accession number: PF01476 Definition: Putative peptidoglycan binding domain Author: Bateman A
			Alignment method of seed: HMM_built_from_alignment
			Source of seed members: Bateman A Gathering cutoffs: 22 22
			Trusted cutoffs: 22.40 22.10 Noise cutoffs: 21.10 21.10
			HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 92324582



			70
Pfam P	rosite	Full Name	Description
			Reference Title: muramidase-2 and Reference Title: Streptococcus faecalis autolysin. Joris B, Englebert S, Chu CP, Kariyama Reference Author: R, Daneo-Moore L, Reference Author: Reference Location: Database Reference: Database reference: Database reference: Database reference: Comment: Tound in a variety Comment: degradation [1]. This Comment: binding function. Number of members: Modular design of the Enterococcus hirae Streptococcus faecalis autolysin. Joris B, Englebert S, Chu CP, Kariyama Shockman GD, Ghuysen JM; FEMS Microbiol Lett 1992;70:257-264. INTERPRO; IPR002482; PFAMB; PB019287; PFAMB; PB040847; PFAMB; PB040977; This domain is about 40 residues long. It is domain may have a general peptidoglycan binding function.
phoslip	I	Phospholipase A2 active sites signatures	Phospholipase A2 (EC 3.1.1.4) (PA2) [1,2] is an enzyme which releases fatty acids from the second carbon group of glycerol. PA2's are small and rigid proteins of 120 amino-acid residues that have four to seven disulfide bonds. PA2 binds a calcium ion which is required for activity. The side chains of two conserved residues, a histidine and an aspartic acid, participate in a 'catalytic network'. Many PA2's have been sequenced from snakes, lizards, bees and mammals. In the latter, there are at least four forms: pancreatic, membrane-associated as well as two less characterized forms. The venom of most snakes contains multiple forms of PA2. Some of them are presynaptic neurotoxins which inhibit neuromuscular transmission by blocking acetylcholine release from the nerve termini. We derived two different signature patterns for PA2's. The first is centered on the active site histidine and contains three cysteines involved in disulfide bonds. The second is centered on the active site aspartic acid and also contains three cysteines involved in disulfide bonds. Description of pattern(s) and/or profile(s) Consensus pattern C-C-x(2)-H-x(2)-C [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL known functional PA2's. However, this pattern will not detect some snake toxins homologous with PA2 but which have lost their catalytic activity as well as otoconin-22, a Xenopus protein from the aragonitic otoconia which is also unlikely to be enzymatically active. Other sequence(s) detected in SWISS-PROT 15. Consensus pattern [LIVMA]-C-{LIVMFYWPCST}-C-D-x(5)-C [D is the active site residue] Sequences are bee PA2, gila monster PA2's, PA2 PL-X from habu and PA2 PA-5 from mulga. Other sequence(s) detected in SWISS-PROT 12. Expert(s) to contact by email Seilhamer JJ. jeff@picryte.com



			9/1
Pfam	Prosite	Full Name	Description
			Last update
			November 1995 / Patterns and text revised. References
			[1]
			Davidson F.F., Dennis E.A.
			J. Mol. Evol. 31:228-238(1990).
			100
			[[2] Gomez F., Vandermeers A., Vandermeers-Piret MC., Herzog R.,
			Rathe J., Stievenart M., Winand J., Christophe J.
			Eur. J. Biochem. 186:23-33(1989).
PI3_PI4_kinase	PDOC00710	Phosphatidylinositol 3-	Phosphatidylinositol 3-kinase (Pl3-kinase) (EC 2.7.1.137) [1] is
		and 4-kinases signatures	an enzyme that phosphorylates phosphoinositides on the 3-hydroxyl group of
			the inositol ring. The exact function of the three products of PI3-kinase -
	:		PI-3-P, PI-3,4-P(2) and PI-3,4,5-P(3) - is not yet known, although it is
			proposed that
			they function as second messengers in cell signalling. Currently, three forms
			of PI3-kinase are known:
			- The mammalian enzyme which is a heterodimer of a 110 Kd
			catalytic chain (p110) and an 85 Kd subunit (p85) which allows it to bind to
			activated tyrosine protein kinases. There are at least two different types
			of p100 subunits (alpha and beta).
			- Yeast TOR1/DRR1 and TOR2/DRR2 [2], PI3-kinases
			required for cell cycle
			activation. Both are proteins of about 280 Kd Yeast VPS34 [3], a PI3-kinase involved in vacuolar sorting and
			segregation.
			VPS34 is a protein of about 100 Kd Arabidopsis thaliana and soybean VPS34 homologs.
			Phosphatidylinositol 4-kinase (PI4-kinase) (EC 2.7.1.67) [4] is
			an enzyme
			that acts on phosphatidylinositol (PI) in the first committed step in the
			production of the second messenger inositol-1,4,5,-
			trisphosphate. Currently
			the following forms of PI4-kinases are known:
			Human DIA kinasa alaha
			- Human Pl4-kinase alpha. - Yeast PIK1, a nuclear protein of 120 Kd.
			- Yeast STT4, a protein of 214 Kd.
			The PI3- and PI4-kinases share a well conserved domain at
			their C-terminal
			section; this domain seems to be distantly related to the catalytic domain of
			protein kinases [2]. We developed two signature patterns from
			the best conserved parts of this domain.
			Four additional proteins belong to this family:
			- Mammalian FKBP-rapamycin associated protein (FRAP) [5], which acts as the
			target for the cell-cycle arrest and immunosuppressive effects of the
			FKBP12-rapamycin complex.
			- Yeast protein ESR1 [6] which is required for cell growth, DNA
!			repair and meiotic recombination.
			- Yeast protein TEL1 which is involved in controlling telomere
			length Yeast hypothetical protein YHR099w, a distantly related
			member of this
	L		family.



			972
Pfam	Prosite	Full Name	Description
			- Fission yeast hypothetical protein SpAC22E12.16C.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIVMFAC]-K-x(1,3)-[DEA]-[DE]-[LIVMC]-R-Q-[DE]-x(4)-Q Sequences known to belong to this class detected by the pattern ALL, except for yeast YHR099w.
			Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [GS]-x-[AV]-x(3)-[LIVM]-x(2)-[FYH]-[LIVM](2)-x-[LIVMF]-x- D-R-H-x(2)-N Sequences known to belong to this class detected by the pattern ALL, except for yeast YHR099w. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Hiles I.D., Otsu M., Volinia S., Fry M.J., Gout I., Dhand R.,
			Panayotou G., Ruiz-Larrea F., Thompson A., Totty N.F., Hsuan J.J., Courtneidge S.A., Parker P.J., Waterfield M.D. Cell 70:419-429(1992). [2] Kunz J., Henriquez R., Schneider U., Deuter-Reinhard M., Movva N., Hall M.N. Cell 73:585-596(1993).
			[3] Schu P.V., Takegawa K., Fry M.J., Stack J.H., Waterfield M.D., Emr S.D. Science 260:88-91(1993).
			[4] Garcia-Bustos J.F., Marini F., Stevenson I., Frei C., Hall M.N. EMBO J. 13:2352-2361(1994).
			[5] Brown E.J., Albers M.W., Shin T.B., Ichikawa K., Keith C.T., Lane W.S., Schreiber S.L. Nature 369:756-758(1994).
			[6] Kato R., Ogawa H. Nucleic Acids Res. 22:3104-3112(1994).
P-II	PDOC00439	P-II protein signatures	The P-II protein (gene glnB) is a bacterial protein important for the control of glutamine synthetase [1,2,3]. In nitrogen-limiting conditions, when the ratio of glutamine to 2-ketoglutarate decreases, P-II is uridylylated on a tyrosine residue to form P-II-UMP. P-II-UMP allows the deadenylation of glutamine synthetase (GS), thus activating the enzyme. Conversely, in nitrogen excess, P-II-UMP is deuridylated and then promotes the adenylation of GS. P-II also indirectly controls the transcription of the GS gene (glnA) by preventing NR-II (ntrB) to phosphorylate NR-I (ntrC) which is the transcriptional activator of glnA. Once P-II is uridylylated, these events are
			P-II is a protein of about 110 amino acid residues extremely well conserved. The tyrosine which is urydylated is located in the central part of the protein.



		9	973
Pfam	Prosite	Full Name	Description In cyanobacteria, P-II seems to be phosphorylated on a serine residue rather than being urydylated. In methanogenic archaebacteria, the nitrogenase iron protein gene (nifH) is
			followed by two open reading frames highly similar to the eubacterial P-II protein [4]. These proteins could be involved in the regulation of nitrogen fixation.
			In the red alga, Porphyra purpurea, there is a glnB homolog encoded in the chloroplast genome.
			Other proteins highly similar to glnB are: - Bacillus subtilis protein nrgB [5]. - Escherichia coli hypothetical protein ybal [6].
			We developed two signature patterns for P-II protein. The first one is a conserved stretch (in eubacteria) of six residues which contains the urydylated tyrosine, the other is derived from a conserved region in the C-terminal part of the P-II protein.
			Description of pattern(s) and/or profile(s)
			Consensus pattern Y-[KR]-G-[AS]-[AE]-Y [The second Y is uridylated] Sequences known to belong to this class detected by the pattern ALL glnB's from eubacteria.
			Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [ST]-x(3)-G-[DY]-G-[KR]-[IV]-[FW]-[LIVM]-x(2)- [LIVM]
·»·			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Magasanik B. Biochimie 71:1005-1012(1989).
			[2] Holtel A., Merrick M. Mol. Gen. Genet. 215:134-138(1988).
			[3] Cheah E., Carr P.D., Suffolk P.M., Vasuvedan S.G., Dixon N.E., Ollis D.L. Structure 2:981-990(1994).
			[4] Sibold L., Henriquet M., Possot O., Aubert JP. Res. Microbiol. 142:5-12(1991).
			[5] Wray L.V. Jr., Atkinson M.R., Fisher S.H. J. Bacteriol. 176:108-114(1994).
			[6] Allikmets R., Gerrard B.C., Court D., Dean M.C. Gene 136:231-236(1993).
pilin	PDOC00342	Prokaryotic N-terminal methylation site	A number of bacteria express filamentous adhesins known as pili. The pili are polar flexible filaments of about 5.4 nm diameter and 2500 nm average length;



	9/4
Pfam Prosite Full Name	Description they consist of a single polypeptide chain (called pilin or fimbrial
	protein) arranged in a helical configuration of five subunits per turn in the
	assembled
	pilus. Gram-negative bacteria produce pilin which are characterized by the
	presence of a very short leader peptide of 6 to 7 residues,
	followed by a methylated N-terminal phenylalanine residue and by a highly
	conserved sequence
	of about 24 hydrophobic residues. This class of pilin is often referred to as
	NMePhe or type-4 pili [1,2].
	Recently a number of bacterial proteins have been sequenced
	which share the following structural characteristics with type-4 pili [3]:
	a) The N-terminal residue, which is methylated, is hydrophobic (generally a
	phenylalanine or a methionine); b) The leader peptide is hydrophilic, consists of 5 to 10 residues
	(with two exceptions, see below) and ends with a glycine;
	c) The fifth residue of the mature sequence is a glutamate which
	seems to be required for the methylation step;
	 d) The first twenty residues of the mature sequence are highly hydrophobic.
	These proteins are listed below:
	Four proteins in an operon involved in a general secretion pathway (GSP)
	for the export of proteins (also called the type II pathway) [4]. These
	proteins have been assigned a different gene name in each of
	the species where they have been sequenced:
	Species Gene names
	Aeromonas hydrophila exeG exeH exel exeJ
	Erwinia chrysanthemi outG outH outI outJ Escherichia coli hofG hofH yheH yheI
	Klebsiella pneumoniae pulG pulH pull pulJ
	Pseudomonase aeruginosa xcpT xcpU xcpV xcpW Vibrio cholerae epsG epsH epsI epsJ
	Xanthomonas campestris xpsG xpsH xpsl xpsJ
	- Vibrio cholerae toxin co-regulated pilin (gene tcpA). This pilin
	has a much longer putative leader peptide (25 residues).
	- Bacillus subtilis comG competence operon proteins 3, 4, and
	5 which are involved for the uptake of DNA by competent Bacillus subtilis
	cells.
	- ppdA, ppdB and ppdC, three Escherichia coli hypothetical proteins found in
	the thyA-recC intergenic region.
	- ppdA, a hypothetical protein near the groeLS operon of Clostridium
	perfringens. The putative leader peptide is 23 residues long.
	We developed a signature pattern based on the N-terminal
	conserved region of all these proteins.
	an aroso protono.
	Description of pattern(s) and/or profile(s)
	Consensus pattern [KRHEQSTAG]-G-[FYLIVM]-[ST]-[LT]-[LIVP]- E-[LIVMFWSTAG](14) [The residue after the G is methylated]

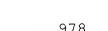


	-	975
Pfam Prosite	Full Name	Description
		Sequences known to belong to this class detected by the pattern
		ALL.
		Other sequence(s) detected in SWISS-PROT NONE. Last update
		November 1995 / Text revised.
		References
		[1]
		Paranchych W., Frost L.S. Adv. Microb. Physiol. 29:53-114(1988).
		7 (dv. 1416) 65. 1 Trycloi. 25.55 114(1555).
		[2]
		Dalrymple B., Mattick J.S.
		J. Mol. Evol. 25:261-269(1987).
		[3]
		Hobbs M., Mattick J.S.
		Mol. Microbiol. 10:233-243(1993).
		[4]
		Salmond G.P.C., Reeves P.J.
		Trends Biochem. Sci. 18:7-12(1993).
PLA2 B	Lysophospholipase	Accession number: PF01735
<u>-</u> -	catalytic domain	Definition: Lysophospholipase catalytic domain
		Author: Bashton M, Bateman A
		Alignment method of seed: Clustalw Source of seed members: Pfam-B_2127 (release 4.1)
		Gathering cutoffs: -283 -283
		Trusted cutoffs: -185.70 -185.70
		Noise cutoffs: -380.50 -380.50
		HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
		Reference Number: [1]
		Reference Medline: 94299545
		Reference Title: Delineation of two functionally distinct domains of
		Reference Title: cytosolic phospholipase A2, a regulatory
		Ca(2+)-dependent
		Reference Title: lipid-binding domain and a Ca(2+)-
		independent catalytic Reference Title: domain.
ļ		Reference Author: Nalefski EA, Sultzman LA, Martin DM,
		Kriz RW, Towler PS,
		Reference Author: Knopf JL, Clark JD; Reference Leasting: LBiol Chem 1994;269:18229 18249
		Reference Location: J Biol Chem 1994;269:18239-18249. Reference Number: [2]
		Reference Medline: 94327513
		Reference Title: The Saccharomyces cerevisiae PLB1 gene
		encodes a protein Reference Title: required for lysophospholipase and
		phospholipase B
		Reference Title: activity.
		Reference Author: Lee KS, Patton JL, Fido M, Hines LK, Kohlwein SD, Paltauf
		Reference Author: F, Henry SA, Levin DE;
		Reference Location: J Biol Chem 1994;269:19725-19730.
		Database Reference: SCOP; 1rlw; fa; [SCOP-USA][CATH-
		PDBSUM] Database Reference INTERPRO; IPR002642;
		Database Reference PDB; 1bci ; 110; 138;
		Database Reference PDB; 1cjy B; 1110; 1430;
		Database Reference PDB; 1cjy A; 110; 498;
		Database Reference PDB; 1rlw; 110; 140; Database Reference PDB; 1cjy B; 1463; 1497;
		Database Reference PDB; 1cjy B; 1539; 1717;
		Database Reference PDB; 1cjy A; 539; 721;
		Comment: This family consists of Lysophospholipase /
		phospholipase B Comment: EC:3.1.1.5 and cytosolic phospholipase A2
		EC:3.1.4 which also
		Comment: has a C2 domain C2.
		Comment: Phospholipase B enzymes catalyse the release of fatty acids from
		release of fatty acids from



		9	76	
Yam	Prosite	Full Name	Description	
				lipids and are capable in vitro of
			nydrolyzing all	o ovtractable form veget cells
		·	Comment: phospholipid [1].	s extractable form yeast cells
				ospholipase A2 associates with
			natural membranes in	oopmonpass : = asses,ates :::
				physiological increases in Ca2+
			and selectively	
			Comment: hydrolyses a	rachidonyl phospholipids [2],
			he aligned region	
			· · · · · · · · · · · · · · · · · · ·	the the carboxy-terminal Ca2+-
			ndependent catalytic	e protein on discussed in [9]
			Comment: domain of th Number of members: 23	e protein as discussed in [2].
		DI AT/I 110 de	Associate numbers DE01477	
PLAT		PLAT/LH2 domain	Accession number: PF01477 Definition: PLAT/LH2 don	
			Author: Bateman A	
	,		Alignment method of seed: Manua	ıl
			Source of seed members: Batem	
			Gathering cutoffs: 25 25	
			Trusted cutoffs: 29.40 29.40	
			Noise cutoffs: -7.90 -7.90	
			HMM build command line: hmmb	
			HMM build command line: hmmo	
				Ilpa; fa; [SCOP-USA][CATH-
			PDBSUM] Database reference: PROSITI	PROFILE; PS50095;
				E_FROFILE, F330033, RO; IPR001024;
				ox; 2; 112;
				pl B; 336; 445;
				pl A; 338; 447;
			Database Reference PDB; 16	th C; 337; 403;
				th A; 339; 405;
			•	th C; 403; 445;
				th A; 405; 447;
	İ			p1 ; 339; 449;
			•	ou8 A; 340; 407;
				ou8 A; 415; 452; pl ; 322; 334;
				a1; 256; 370;
			-	m6 A; 256; 370;
				m6 B; 256; 370;
				md A; 256; 370;
				ımd B; 256; 370;
				is found in a variety of
			membrane or	
			Comment: lipid associa PLAT	ted proteins. It is called the
	}			1, Lipoxygenase, Alpha-Toxin)
			domain or	genase homology) domain. The
			Comment: LH2 (Lipoxy known structure	genase nomology) domain. Th
				c lipase shows this domain
			binds to procolipase	
				hich mediates membrane
			association.	e appoints that this demain
			Comment: So it appea mediates membrane	s possible that this domain
				via other protein binding
			partners. The	this domain is known for many
			Comment: structure of members of the	una domain la khown for many
			Comment: family and i Number of members: 82	s composed of a beta sandwich
	1	Potato leaf roll virus	Accession number: PF01696)
PLRV ORF5		readthrough protein		I virus readthrough protein
PLRV_ORF5		i caatii oagii protoii		
PLRV_ORF5		readthrough protein	Author: Bashton M, Ba	ateman A
PLRV_ORF5		Toddinodgii protein	Author: Bashton M, Bas	ateman A alw
PLRV_ORF5		Teautinough protein	Author: Bashton M, Ba Alignment method of seed: Clusta Source of seed members: Pfam	ateman A alw
PLRV_ORF5		Todal Hough protest	Author: Bashton M, Bas	ateman A alw B_1335 (release 4.1)

Pfam	Prosite	Full Name	Description
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			Reference Number: [1] Reference Medline: 94233771
			Reference Title: Changes in the amino acid sequence of the
			coat protein
			Reference Title: readthrough domain of potato leafroll
			luteovirus affect the Reference Title: formation of an epitope and aphid
			Heference Title: formation of an epitope and aphid transmission.
		1	Reference Author: Jolly CA, Mayo MA;
			Reference Location: Virology 1994;201:182-185.
			Database Reference INTERPRO; IPR002929;
			Comment: This family consists mainly of the potato lea
			roll virus Comment: readthrough protein. This is generated via a
			readthrough
			Comment: of open reading frame 3 a coat protein
			allowing transcription
			Comment: of open reading frame 5 to give an extender
			coat protein Comment: with a large c-terminal addition or read
			Comment: with a large c-terminal addition or read through domain [1].
			Comment: The readthrough protein is thought to play a
i			role in the
			Comment: circulative aphid transmission of potato leaf
			roll virus [1]. Comment: Also in the family is open reading frame 6
			Comment: Also in the family is open reading frame 6 from beet western
			Comment: yellows virus and potato leaf roll virus both
			luteovirus and
			Comment: an unknown protein from cucurbit aphid-
			borne yellows virus a Comment: closterovirus.
			Number of members: 28
PMSR		Peptide methionine	Accession number: PF01625
		sulfoxide reductase	Definition: Peptide methionine sulfoxide reductase
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_1111 (release 4.1) Gathering cutoffs: -62 -62
			Gathering cutoffs: -62 -62 Trusted cutoffs: -28.00 -28.00
			Noise cutoffs: -96.70 -96.70
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1] Reference Medline: 96353931
			Reference Medline: 96353931 Reference Title: Peptide methionine sulfoxide reductase
			contributes to the
			Reference Title: maintenance of adhesins in three major
			pathogens.
			Reference Author: Wizemann TM, Moskovitz J, Pearce BJ, Cundell D, Arvidson
			Reference Author: CG, So M, Weissbach H, Brot N, Masure
			HR;
			Reference Location: Proc Natl Acad Sci USA 1996;93:7985-7990.
			Reference Number: [2]
			Reference Medline: 96312545
ļ			Reference Title: Cloning the expression of a mammalian gene involved in the
	j		Reference Title: reduction of methionine sulfoxide residues
			in proteins.
			Reference Author: Moskovitz J, Weissbach H, Brot N;
			Reference Location: Proc Natl Acad Sci U S A 1996;93:2095-
	·		Reference Location: Proc Natl Acad Sci U S A 1996;93:2095-2099.
	·		Reference Location: 2099. Database Reference Comment: Proc Natl Acad Sci U S A 1996;93:2095- INTERPRO; IPR002569; This enzyme repairs damaged proteins.
			Reference Location: Proc Natl Acad Sci U S A 1996;93:2095-2099. Database Reference Comment: INTERPRO; IPR002569; This enzyme repairs damaged proteins. Methionine sulfoxide in proteins
			Reference Location: 2099. Database Reference Comment: INTERPRO; IPR002569; This enzyme repairs damaged proteins. Methionine sulfoxide in proteins Comment: is reduced to methionine.
ollen allerg 2			Reference Location: Proc Natl Acad Sci U S A 1996;93:2095-2099. Database Reference Comment: INTERPRO; IPR002569; This enzyme repairs damaged proteins. Methionine sulfoxide in proteins



		S	978
Pfam	Prosite	Full Name	Description
		allergen)	Definition: Ribonuclease (pollen allergen)
			Author: Bateman A
			Alignment method of seed: Clustalw Source of seed members: Pfam-B_1050 (release 4.1)
			Gathering cutoffs: -3 -3
			Trusted cutoffs: 23.10 23.10
			Noise cutoffs: -29.40 -29.40
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 95246885
			Reference Title: Major allergen PhI p Vb in timothy grass is a
			Inovel pollen Reference Title: RNase.
			Reference Author: Bufe A, Schramm G, Keown MB, Schlaak
			M, Becker WM;
			Reference Location: Febs lett 1995;363:6-12.
		i	Database Reference INTERPRO; IPR002914;
			Database reference: PFAMB; PB037130;
			Comment: This family contains grass pollen proteins of
			group V. Comment: Swiss:Q40963 has been shown to possess
			ribonuclease
			Comment: activity [1].
			Number of members: 27
DOD N		Pyruvate	Accession number: PF01855
POR_N		flavodoxin/ferredoxin	Definition: Pyruvate flavodoxin/ferredoxin oxidoreductase
		oxidoreductase (N	(N terminus)
		terminus)	Author: Bateman A
		,	Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_323 (release 4.2)
			Gathering cutoffs: -116 -116
			Trusted cutoffs: -113.60 -113.60 Noise cutoffs: -119.50 -119.50
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 96125254
			Reference Title: Molecular and phylogenetic
			characterization of pyruvate and Reference Title: 2-ketoisovalerate ferredoxin
			oxidoreductases from
			Reference Title: Pyrococcus furiosus and pyruvate
			ferredoxin oxidoreductase
			Reference Title: from Thermotoga maritima.
			Reference Author: Kletzin A, Adams MW;
			Reference Location: J Bacteriol 1996;178:248-257. Reference Number: [2]
			Reference Medline: 94022264
			Reference Title: Growth of the cyanobacterium Anabaena
			on molecular
			Reference Title: nitrogen: NifJ is required when iron is
			limited.
[Reference Author: Bauer CC, Scappino L, Haselkorn R; Reference Location: Proc Natl Acad Sci U S A 1993;90:8812-
			Reference Location: Proc Natl Acad Sci U S A 1993;90:8812-8816.
			Reference Number: [3]
			Reference Medline: 99140300
			Reference Title: Crystal structures of the key anaerobic
	1		enzyme
			Reference Title: pyruvate:ferredoxin oxidoreductase, free
			and in complex Reference Title: with pyruvate.
			Reference Author: Chabriere E, Charon MH, Volbeda A,
			Pieutle L, Hatchikian
			Reference Author: EC, Fontecilla-Camps JC;
			Reference Location: Nat Struct Biol 1999;6:182-190.
			Database Reference: SCOP; 2pda; fa; [SCOP-USA][CATH-
			PDBSUM] Database Reference: SCOP; 2pda; fa; [SCOP-USA][CATH-
·			Database Reference: SCOP; 2pda; fa; [SCOP-USA][CATH-PDBSUM]
			Database Reference INTERPRO; IPR002880;
	1		Database Reference PDB; 1b0p A; 43; 328;
L			

			979	
Pfam	Prosite	Full Name	Description	
			Database Reference	, (, , ,
			Database Reference	, , , , , , , , , , , , , , , , , , , ,
			Database Reference: Database reference:	,-,-,-,-,,-,,
			Comment:	This family includes the N terminal region of
			the pyruvate ferredo	
			Comment:	oxidoreductase, corresponding to the first
			two structural domai	
			Comment:	This region is involved in inter subunit
			contacts [3]. Pyruvat	
			Comment: step in the fermentat	oxidoreductase (POR) catalyses the final
			Comment:	of carbohydrates in anaerobic
			microorganisms [1].	
			Comment:	oxidative decarboxylation of pyruvate with
			the participation of	, , , ,
			Comment:	thiamine followed by the transfer of an
			acetyl moiety to coer	
			Comment:	A for the synthesis of acetyl-CoA [1]. The
			family also includes Comment:	pyruvate flavodoxin oxidoreductase as
			encoded by the nifJ	
			Comment:	cyanobacterium which is required for growth
			on molecular nitroger	
			Comment:	when iron is limited [2].
			Number of members	: 55
PPE		PPE family	Accession number:	PF00823
		1 1 L ranning		PE family
				ateman A
			Alignment method of	seed: Clustalw_manual
			Source of seed mem	bers: Pfam-B_297 (release 3.0)
			Gathering cutoffs:	-90 -90
			Trusted cutoffs: Noise cutoffs:	-88.20 -88.20 105.20 105.20
				-105.30 -105.30 I line: hmmbuild -F HMM SEED
			HMM build command	I line: hmmcalibrateseed 0 HMM
			Reference Number:	[1]
			Reference Medline:	98295987
			Reference Title:	Deciphering the biology of Mycobacterium
			tuberculosis from	Aba associate socialistics
			Reference Title: Reference Author:	the complete genome sequence.
			Reference Location:	Nature 1998;393:537-544.
			Database Reference	INTERPRO; IPR000030;
			Database reference:	PFAMB; PB040834;
			Comment:	This family named after a PPE motif near to
			the amino	towninus of the descrip. The DDE feeth, of
			Comment: proteins	terminus of the domain. The PPE family of
			Comment:	all contain an amino-terminal region of
			about 180	
			Comment:	amino acids. The carboxyl terminus of this
			family	
			Comment:	are variable, and on the basis of this region
			fall Comment:	into at least three groups. The MOTO
			subgroup has	into at least three groups. The MPTR
			Comment:	tandem copies of a motif NXGXGNXG. The
			second subgroup	,
			Comment:	contains a conserved motif at about position
			350.	The third annual control of the cont
	i		Comment: terminal	The third group are only related in the amino
			comment:	region.
			Comment:	The function of these proteins is uncertain
			but it	and a mode proteins is uncertain
			Comment:	has been suggested that they may be
	1		related to	
			Comment:	antigenic variation of Mycobacterium
			tuberculosis [1]. Number of members:	75
			Number of members:	75
PRA-CH		Phosphoribosyl-AMP	Accession number:	PF01502
		Jopinon booy :- Alvii	TOOCOOTOR HUITIDEL.	1101006





Pfam	Prosite	Full Name	Description
Pfam	Prosite		Description Definition: Phosphoribosyl-AMP cyclohydrolase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_782 (release 4.0) Gathering cutoffs: 25 25 Trusted cutoffs: 88.20 88.20 Noise cutoffs: -44.30 -44.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 99129952 Reference Title: word of a Reference Title: cyclohydrolase: purification and cyclohydrolase: purification
			histidine Comment: biosynthetic pathway. It requires Zn ions for activity. Number of members: 28
PRA-PH		Phosphoribosyl-ATP pyrophosphohydrolase	Accession number: PF01503 Definition: Phosphoribosyl-ATP pyrophosphohydrolase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_784 (release 4.0) Gathering cutoffs: 6 6 Trusted cutoffs: 12.10 12.10 Noise cutoffs: 1.00 1.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 79216449 Reference Title: The product of the his4 gene cluster in Saccharomyces Reference Author: Acesey JK Jr, Bigelis R, Fink GR; Reference Author: J Biol Chem 1979 Aug 10;254:7427-7433. Reference Number: [2] Reference Number: [2] Reference Title: Reference Title: HIS4 gene product of Saccharomyces cerevisiae and the hisIE Reference Title: and hisD gene products of Escherichia coli Reference Title: Typhimurium. Bruni CB, Carlomagno MS, Formisano S, Mol Gen Genet 1986;203:389-396. INTERPRO; IPR002497; This enzyme catalyses the second step in the histidine Comment: biosynthetic pathway. Number of members: 32
PseudoU_synth_1		tRNA pseudouridine synthase	Accession number: PF01416 Definition: tRNA pseudouridine synthase Previous Pfam IDs: PseudoU_synt; Author: Howe K Alignment method of seed: Clustalw Source of seed members: swissprot Gathering cutoffs: 30 30 Trusted cutoffs: 39.10 39.10 Noise cutoffs: -55.00 -55.00 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 98254513



	1		Description
Pfam	Prosite	Full Name	Description Reference Title: Transfer RNA-pseudouridine synthetase
			Pus1 of Saccaromyces
			Reference Title: cerevisiae contains one atom of zinc
	İ		essential for its
			Reference Title: native conformation and tRNA recognition.
			Reference Author: Arluison V, Hountondji C, Robert B,
			Grosjean H; Reference Location: Biochemistry 1998;37:7268-7276.
			Database Reference INTERPRO; IPR001406;
			Database reference: PFAMB; PB027500;
			Comment: Involved in the formation of pseudouridine at
			the anticodon stem
			Comment: and loop of transfer-RNAs
			Comment: Pseudouridine is an isomer of uridine (5-
			(beta-D-ribofuranosyl) Comment: uracil, and id the most abundant modified
			nucleoside found in
			Comment: all cellular RNAs.
			Comment: The TruA-like proteins also exhibit a
			conserved sequence with
			Comment: a strictly conserved aspartic acid, likely
			involved in catalysis
			Number of members: 31
PseudoU synth 2		RNA pseudouridylate	Accession number: PF00849
i aeudoo_ayniii_2		synthase	Definition: RNA pseudouridylate synthase
			Previous Pfam IDs: YABO;
			Author: Bateman A
	i		Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_421 (release 3.0)
			Gathering cutoffs: 20 20
			Trusted cutoffs: 20.90 20.90
			Noise cutoffs: -44.40 -44.40
			HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 96079974
			Reference Title: A dual-specificity pseudouridine synthase:
			an Escherichia
			Reference Title: coli synthase purified and cloned on the
			basis of its
			Reference Title: specificity for psi 746 in 23S RNA is also
			specific for psi Reference Title: 32 in tRNA(phe).
			Reference Author: Wrzesinski J, Nurse K, Bakin A, Lane BG,
			Ofengand J;
			Reference Location: RNA 1995;1:437-448.
			Database Reference: PROSITE; PDOC00869
			Database Reference: PROSITE; PDOC00885
			Database Reference INTERPRO; IPR000613;
			Database reference: PFAMB; PB041160;
			Database reference: PFAMB; PB041232;
			Comment: Members of this family are involved in modifying bases in RNA molecules.
			Comment: They carry out the conversion of uracil
			bases to pseudouridine. This family
			Comment: includes RluD Swiss:P33643, a
			pseudouridylate synthase that converts
	}		Comment: specific uracils to pseudouridine in 23S
			rRNA. RluA from E. coli
			Comment: converts bases in both rRNA and tRNA [1].
			Number of members: 78
PWI		PWI domain	Accession number: PF01480
1 VV I		T TYL GOLLIAILI	Definition: PWI domain
			Author: Bateman A
			Alignment method of seed: Clustalw_manual
			Source of seed members: [1]
			Gathering cutoffs: 25 25
			Trusted cutoffs: 64.40 64.40
			Noise cutoffs: -3.50 -3.50
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
	I	L	Reference Number: [1]

Pfam	Prosite	Full Name	Poporiation
+ (411)	FIUSILE	Full Name	Description Reference Medline: 10322432
			Reference Title: The PWI motif: a new protein domain in splicing factors.
			Reference Author: Blencowe BJ, Ouzounis CA:
			Reference Location: Trends Biochem Sci 1999;24:179-180.
			Database Reference INTERPRO; IPR002483;
			Number of members: 11
R3H		R3H domain	Accession number: PF01424
		, , , , , , , , , , , , , , , , , , , ,	Definition: R3H domain
			Author: Bateman A
			Alignment method of seed: Manual
			Source of seed members: Medline:99003905
			Gathering cutoffs: 25 25
			Trusted cutoffs: 59.30 59.30 Noise cutoffs: 5,10 5,10
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
		1	Reference Medline: 99003905
			Reference Title: The R3H motif: a domain that binds single-
			stranded nucleic
			Reference Title: acids.
			Reference Author: Grishin NV; Reference Location: Trends Biochem Sci 1998:23:329-330
			Reference Location: Trends Biochem Sci 1998;23:329-330. Database Reference INTERPRO; IPR001374;
			Database reference: PFAMB; PB041444;
•			Comment: The name of the R3H domain comes from
			the characteristic spacing
		Comment: of the most conserved arginine and histidine	
			residues. The
			Comment: function of the domain is predicted to be
			binding ssDNA. Number of members: 28
			Number of members: 28
RepB_protein		Initiator RepB protein	Accession number: PF01051
			Definition: Initiator RepB protein
			Author: Finn RD, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_313 (release 3.0) Gathering cutoffs: 14 14
			Trusted cutoffs: 19.00 16.20
			Noise cutoffs: 11.80 12.90
			HMM build command line: hmmbuild -f HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 98284148
			Reference Title: Replication and control of circular bacterial plasmids.
			Reference Author: del Solar G, Giraldo R, Ruiz-Echevarria
			MJ, Espinosa M,
			Reference Author: Diaz-Orejas R;
			Reference Location: Microbiol Mol Biol Rev 1998;62:434-464.
			Reference Number: [2]
			Reference Medline: 97324207 Reference Title: Initiation of replication of plasmid pMV158
			Reference Title: Initiation of replication of plasmid pMV158: mechanisms of
			Reference Title: DNA strand-transfer reactions mediated by
			the initiator
			Reference Title: RepB protein.
			Reference Author: Moscoso M, Eritja R, Espinosa M;
			Reference Location: J Mol Biol 1997;268:840-856.
	[Database Reference INTERPRO; IPR000525;
			Database Reference PDB; 1rep C; 198; 240;
			Database reference: PFAMB; PB000509; Comment: This protein is an initiator of plasmid
			Comment: This protein is an initiator of plasmid replication.
			Comment: RepB possesses nicking-closing
			(topoisomerase I) like activity.
			Comment: It is also able to perform a strand transfer
	1		reaction on ssDNA
			Comment: that contains its target.

⊃fam	Prosite	Full Name	Description
Rhomboid		Rhomboid family	Accession number: PF01694
			Definition: Rhomboid family
			Author: Sohrmann M, Bateman A
			Alignment method of seed: Clustalw
		İ	Source of seed members: Pfam-B 1399 (release 4.1)
			Gathering cutoffs: 25 25
			Trusted cutoffs: 143.60 143.60
			Noise cutoffs: -43.60 -43.60 HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 90249726
			Reference Title: rhomboid, a gene required for dorsoventral
			axis
			Reference Title: establishment and peripheral nervous
			system development in
	!		Reference Title: Drosophila melanogaster.
			Reference Author: Bier E, Jan LY, Jan YN;
			Reference Location: Genes Dev 1990;4:190-203.
			Database Reference INTERPRO; IPR002610;
			Database reference: PFAMB; PB041113;
			Comment: This family contains integral membrane
			proteins that are
			Comment: related to Drosophila rhomboid protein
			Swiss:P20350. Members
			Comment: of this family are found in bacteria and
			eukarvotes. These
			Comment: proteins contain three strongly conserved
			histidines in the
			Comment: putative transmembrane regions that may
			be involved in the
			Comment: as yet unknown function of these proteins.
		}	Number of members: 27
Ribosomal L18ae		Ribosomal L18ae protein	Accession number: PF01775
_		family	Definition: Ribosomal L18ae protein family
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: PSI-BLAST Q02543
			Gathering cutoffs: 25 25 Trusted cutoffs: 136.70 136.70
			Trusted cutoffs: 136.70 136.70 Noise cutoffs: -99.80 -99.80
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Database Reference INTERPRO; IPR002670;
			Number of members: 11
			Namber of members.
Dil consol I 015	PDOC00899	Ribosomal protein L21	Ribosomal protein L21 is one of the proteins from the large
Ribosomal_L21p	PD0000099	signature	ribosomal subunit.
		signature	In Escherichia coli, L21 is known to bind to the 23S rRNA in the
			presence of
	•		L20. It belongs to a family of ribosomal proteins which, on the
		1	basis of
			sequence similarities, groups:
			, 5
			- Eubacterial L21.
			- Marchantia polymorpha chloroplast L21.
			- Cyanelle L21.
			- Spinach chloroplast L21 (nuclear-encoded).
			Eubacterial L21 is a protein of about 100 amino-acid residues, the
			mature form
			of the spinach chloroplast L21 has 200 residues. As a signature
			pattern, we
			selected a conserved region located in the C-terminal section
			of these
			proteins.
			Description of pattern(s) and/or profile(s)
1			THE P
	1	!	Consensus pattern [IVT]-x(3)-[KR]-x(3)-[KRQ]-K-x(6)-G-[HF]-R-



am i	Prosite	Full Name	Description description
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised.
ibosomal L22e	1	Ribosomal L22e protein family	Accession number: PF01776 Definition: Ribosomal L22e protein family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P56628 Gathering cutoffs: 25 25 Trusted cutoffs: 262.80 262.80 Noise cutoffs: -52.00 -52.00 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference INTERPRO; IPR002671; Number of members: 11
libosomal_L27e		Ribosomal L27e protein family	Accession number: PF01777 Definition: Ribosomal L27e protein family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P51419 Gathering cutoffs: 25 25 Trusted cutoffs: 326.90 326.90 Noise cutoffs: -47.80 -47.80 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference INTERPRO; IPR001141; Number of members: 9
Ribosomal_L29	PDOC00501	Ribosomal protein L29 signature	Ribosomal protein L29 is one of the proteins from the large ribosomal subunit. L29 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L29 Red algal L29 Archaebacterial L29 Mammalian L35 - Caenorhabditis elegans L35 (ZK652.4) Yeast L35. L29 is a protein of 63 to 138 amino-acid residues. As a signature pattern, we selected a conserved region located in the central section of L29
			Description of pattern(s) and/or profile(s) Consensus pattern [KNQS]-[PSTLN]-x(2)-[LIMFA]-[KRGSAN]-x-[LIVYSTA]-[KR]- [KRHQS]-[DESTANRL]-[LIV]-A-[KRCQVT]-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 2. Last update December 1999 / Pattern and text revised. References [1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).
Ribosomal_L31e	PDOC0088	Ribosomal protein L316 signature	A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian L31 [1] Chlamydomonas reinhardtii L31.

			985
Pfam	Prosite	Full Name	Description
			- Halobacterium marismortui HL30 [2].
			These proteins have 87 to 128 amino-acid residues. As a
			signature pattern, we
			selected a conserved region located in the central section.
			Description of pattern(s) and/or profile(s)
			Consensus pattern V-[KR]-[LIVM]-x(3)-[LIVM]-N-x-[AKH]-x-W-x-
	1		[KR]-G
			Sequences known to belong to this class detected by the pattern ALL.
	1		Other sequence(s) detected in SWISS-PROT NONE.
			Last update
	İ		July 1999 / Pattern and text revised. References
			[1]
	İ		Tanaka T., Kuwano Y., Kuzumaki T., Ishikawa K., Ogata K.
			Eur. J. Biochem. 162:45-48(1987).
			[2]
			Bergmann U., Arndt E.
			Biochim. Biophys. Acta 1050:56-60(1990).
Ribosomal_L35Ae	PDOC00849	Ribosomal protein L35Ae	A number of eukaryotic and archaebacterial ribosomal proteins
		signature	can be grouped
			on the basis of sequence similarities. One of these families consists of:
			COLORGE UI.
			- Vertebrate L35A.
			- Caenorhabditis elegans L35A (F10E7.7). - Yeast L37A/L37B (Rp47).
			- Yeast L3/A/L3/B (Hp4/) Pyrococcus woesei L35A homolog [1].
			These proteins have 87 to 110 amino-acid residues. As a
			signature pattern, we selected a highly conserved stretch of 22 residues in the C-
			terminal part of
			these proteins.
			Description of pattern(s) and/or profile(s)
			Consensus pattern G-K-[LIVM]-x-R-x-H-G-x(2)-G-x-V-x-A-x-F-
			x(3)-[LI]-P
			Sequences known to belong to this class detected by the pattern
			Other sequence(s) detected in SWISS-PROT NONE.
			Last update
		u l	November 1997 / Pattern and text revised. References
			[1]
			Ouzounis C., Kyrpides N., Sander C.
			Nucleic Acids Res. 23:565-570(1995).
Ribosomal_L35p	PDOC00721	Ribosomal protein L35	Ribosomal protein L35 is one of the proteins from the large
		signature	subunit of the
			ribosome. It belongs to a family of ribosomal proteins which, on the basis of
	i		sequence similarities [1], groups:
			- Eubacterial L35.
			Plant chloroplast L35 (nuclear-encoded).Red algal chloroplast L35.
			- Cyanelle L35.
			LSE in a basic protein of CO to TO and
			L35 is a basic protein of 60 to 70 amino-acid residues. As a signature pattern
			we selected a conserved region in the N-terminal section.
			_



		,	86
Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIVM]-K-[TV]-x(2)-[GSA]-[SAILV]-x-K-R-
			[LIVMFY]-[KRLS] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update
		ı	December 1999 / Pattern and text revised. References
			[1]
			Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).
Ribosomal_L36e	PDOC00916	Ribosomal protein L36e	A number of eukaryotic ribosomal proteins can be grouped on
_		signature	the basis of sequence similarities. One of these families consists of:
			- Mammalian L36 [1]. - Drosophila L36 (M(1)1B).
			- Caenorhabditis elegans L36 (F37C12.4). - Candida albicans L39.
			- Yeast YL39.
			These proteins have 99 to 104 amino acids. As a signature
			pattern, we selected a conserved region in the central part of these proteins.
			Selected a conserved region in the contrar part of those proteins.
			Description of pattern(s) and/or profile(s)
			Consensus pattern P-Y-E-[KR]-R-x-[LIVM]-[DE]-[LIVM](2)-[KR]
			Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Last update November 1997 / First entry.
			References
			[1] Chan YL., Paz V., Olvera J., Wool I.G.
			Biochem. Biophys. Res. Commun. 192:849-853(1993).
Ribosomal_L37ae		Ribosomal L37ae protein	
		family	Author: Bateman A
			Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P54051
			Gathering cutoffs: 25 25
			Trusted cutoffs: 145.10 145.10
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Database Reference INTERPRO; IPR002674;
			Comment: This ribosomal protein is found in
			archaebacteria and Comment: eukaryotes. It contains four conserved
			cysteine
			Comment: residues that may bind to zinc. Number of members: 15
Ribosomal_L37e	PDOC00827	Ribosomal protein L37e	
		signature	can be grouped on the basis of sequence similarities. One of these families consists of:
			- Mammalian L37 [1].
			- Leishmania infantum L37 [2].
			- Fission yeast YL35 [3]. - Halobacterium marismortui L37e (L35e) [4].
			These proteins have 56 to 96 amino-acid residues. As a
			signature pattern, we

Pfam	D	Full Name	98 /
ran.	Prosite	rui Name	Description selected a highly conserved region located in the N-terminal part of these proteins.
			Description of pattern(s) and/or profile(s) Consensus pattern G-T-x-[SA]-x-G-x-[KR]-x(3)-[STLR]-x(0,1)-H-x(2)-C-x-R-C-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Chan YL., Paz V., Olvera J., Wool I.G. Biochem. Biophys. Res. Commun. 192:590-596(1993). [2] Myler P.J., Tripp C.A., Thomas L., Venkataraman G.M., Merlin G., Stuart K. Mol. Biochem. Parasitol. 62:147-152(1993).
			[3] Otaka E., Higo KI., Itoh T. Mol. Gen. Genet. 191:519-524(1983). [4] Bergmann U., Wittmann-Liebold B. Biochim. Biophys. Acta 1173:195-200(1993).
Ribosomal_L38e		Ribosomal L38e protein family	Accession number: PF01781 Definition: Ribosomal L38e protein family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P23411 Gathering cutoffs: 25 25 Trusted cutoffs: 127.60 127.60 Noise cutoffs: -24.50 -24.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 91207349 Reference Title: The primary structure of rat ribosomal protein L38. Reference Author: Kuwano Y, Olvera J, Wool IG; Reference Location: Biochem Biophys Res Commun 1991;175:551-555. Database Reference Number of members: 8
Ribosomal_L39	PDOC00050	Ribosomal protein L39e signature	A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian L39 [1] Plants L39 Yeast L46 [2] Archebacterial L39e [3]. These proteins are very basic. About 50 residues long, they are the smallest proteins of eukaryotic-type ribosomes. As a signature pattern, we selected a conserved region in the C-terminal section of these proteins.
			Description of pattern(s) and/or profile(s) Consensus pattern [KRA]-T-x(3)-[LIVM]-[KRQF]-x-[NHS]-x(3)-R-



	B		Description
Pfam	Prosite	Full Name	Description [NHY]-W-R-R
			Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update
			July 1998 / Pattern and text revised. References
			Lin A., McNally J., Wool I.G.
			J. Biol. Chem. 259:487-490(1984).
			[2]
			Leer R.J., van Raamsdonk-Duin M.M.C., Kraakman P., Mager W.H., Planta R.J.
			Nucleic Acids Res. 13:701-709(1985).
			[3]
			Ramirez C., Louie K.A., Matheson A.T.
			FEBS Lett. 250:416-418(1989).
Ribosomal_L4	PDOC00724	Ribosomal protein L1e	A number of eukaryotic and archaebacterial ribosomal proteins
Tibosoma_L+	, 50000121	signature	can be grouped
			on the basis of sequence similarities. One of these families consists [1,2,3,
			4] of:
			- Vertebrate L1 (L4).
			- Drosophila L1.
			- Plant L1.
			- Yeast L2 (Rp2). - Fission yeast L2.
			- Halobacterium marismortui HmaL4 (HL6).
			- Methanococcus jannaschii MJ0177.
			These proteins have 246 (archaebacteria) to 427 (human)
			amino acids. As a signature pattern, we selected a conserved region in the N-
			terminal part of
			these proteins.
			Description of pattern(s) and/or profile(s)
			Consensus pattern N-x(3)-[KRM]-x(2)-A-[LIVT]-x-S-A-[LIV]-x-A-
			ST]-[SGA]- x(7)-[RK]-[GS]-H Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
	1		Last update November 1997 / Pattern and text revised.
			References
			[1] Rafti F., Gargiulo G., Manzi A., Malva C., Graziani F.
			Nucleic Acids Res. 17:456-456(1989).
			[2]
			Presutti C., Villa T., Bozzoni I.
			Nucleic Acids Res. 21:3900-3900(1993).
			[3]
			Bagni C., Mariottini P., Annesi F., Amaldi F. Arndt E., Kroemer W., Hatakevama T.
			Biochim. Biophys. Acta 1216:475-478(1993). J. Biol. Chem.
			265:3034-3039(1990).
Ribosomal S20p	-	Ribosomal protein S20	Accession number: PF01649
		·	Definition: Ribosomal protein S20
			Author: Bateman A Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_1685 (release 4.1)
	I .		
			Gathering cutoffs: 25 25 Trusted cutoffs: 57.30 57.30

		ğ	989
Pfam	Prosite	Full Name	Description
			HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 88230452 Reference Title: 88230452 Interaction of proteins S16, S17 and S20 with 16 S Reference Title: ribosomal RNA. Reference Author: Stern S, Changchien LM, Craven GR, Noller HF; Reference Location: J Mol Biol 1988;200:291-299. INTERPRO; IPR002583; Bacterial ribosomal protein S20 interacts with 16S rRNA [1]. Number of members: 29
Ribosomal_S27e	PDOC00898	Ribosomal protein S27e signature	A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of [1]:
			- Mammalian S27 (human S27 was originally known as metallopan-stimulin 1) Chlamydomonas reinhardtii S27 Entamoeba histolytica S27 Yeast S27 Archaebacterial S27e. These proteins have from 62 to 87 amino acids. They contain, in their central section, a putative zinc-finger region of the type C-x(2)-C-x(14)-C-x(2)-C. We have selected that region as a signature pattern.
			Description of pattern(s) and/or profile(s) Consensus pattern [QKT]-C-x(2)-C-x(6)-F-[GSD]-x-[PSA]-x(5)-C-x(2)-C-[GSA]-x(2)-[LV]-x(2)-P-x-G [The four C's are potential zinc ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Chan YL., Suzuki K., Olvera J., Wool I.G. Nucleic Acids Res. 21:649-655(1993).
Ribosomal_S3_C	PDOC00474	Ribosomal protein S3 signature	Ribosomal protein S3 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial S3 Algal and plant chloroplast S3 Cyanelle S3 Plant mitochondrial S3 Plant mitochondrial S3 Vertebrate S3 Insect S3 Caenorhabditis elegans S3 (C23G10.3) Yeast S3 (Rp13). S3 is a protein of 209 to 559 amino-acid residues. As signature patterns, we selected a conserved region located in the C-terminal section.



	r		990
Pfam	Prosite	Full Name	Description Description of pattern(s) and/or profile(s)
			Consensus pattern [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]- x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS] Sequences known to belong to this class detected by the pattern
			ALL, except for some mitochondrial S3. Other sequence(s) detected in SWISS-PROT NONE.
			Expert(s) to contact by email Hallick R.B. hallick@arizona.edu
·			Last update December 1999 / Pattern and text revised.
			References [1]
			Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).
Ribosomal_S3_N	PDOC00474	Ribosomal protein S3 signature	Ribosomal protein S3 is one of the proteins from the small ribosomal subunit.
			In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on
			the basis of sequence similarities [1], groups:
			- Eubacterial S3. - Algal and plant chloroplast S3. - Cyanelle S3.
			- Archaebacterial S3. - Plant mitochondrial S3.
			Vertebrate S3.Insect S3.Caenorhabditis elegans S3 (C23G10.3).Yeast S3 (Rp13).
			S3 is a protein of 209 to 559 amino-acid residues. As signature patterns, we selected a conserved region located in the C-terminal section.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]- x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS] Sequences known to belong to this class detected by the pattern ALL, except for some mitochondrial S3. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Hallick R.B. hallick@arizona.edu
			Last update December 1999 / Pattern and text revised. References
			[1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).
RimM		RimM	Accession number: PF01782 Definition: RimM Author: Bateman A Alignment method of seed: Clustalw
			Source of seed members: PSI-BLAST P51419 Gathering cutoffs: 25 25
			Trusted cutoffs: 49.00 49.00 Noise cutoffs: -66.10 -66.10 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1] Reference Medline: 98083058 Reference Title: RimM and RbfA are essential for efficient
			processing of 16S Reference Title: rRNA in Escherichia coli. Reference Author: Bylund GO, Wipemo LC, Lundberg LA,

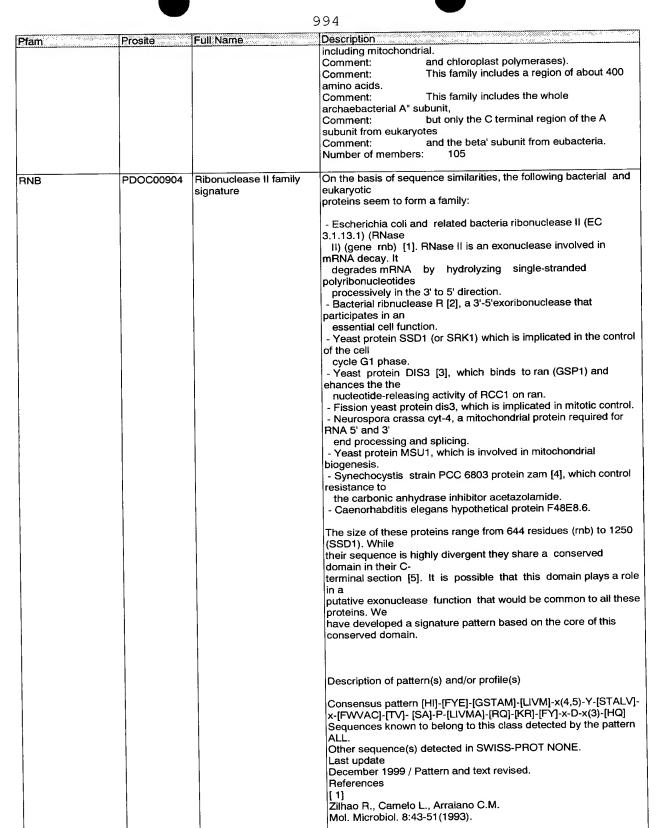


			991
Pfam	Prosite	Full Name	Description Wikstrom PM;
			Reference Location: J Bacteriol 1998;180:73-82.
			Database Reference INTERPRO: IPR002676;
			Comment: The RimM protein is essential for efficient
			processing of 16S rBNA [1].
•			Comment: The RimM protein was shown to have
		1	affinity for free ribosomal 30S
			Comment: subunits but not for 30S subunits in the 70S
			ribosomes [1].
			Number of members: 14
DAIA de DAIA del		RNA dependent RNA	Accession number: PF00680
RNA_dep_RNA_pol		polymerase	Definition: RNA dependent RNA polymerase
		1	Author: Bateman A
		1	Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_32 (release 2.1) Gathering cutoffs: -127 -127
			Gathering cutoffs: -127 -127 Trusted cutoffs: -117.00 -117.00
			Noise cutoffs: -137.30 -137.30
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Database Reference: SCOP; 1rdr; fa; [SCOP-USA][CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR001205;
		ļ	Database Reference PDB; 1rdr; 12; 37;
		1	Database Reference PDB; 1rdr; 182; 460;
			Database Reference PDB; 1rdr; 67; 97;
			Database reference: PFAMB; PB039844;
			Database reference: PFAMB; PB040630;
			Database reference: PFAMB; PB040631;
			Database reference: PFAMB; PB040844; Database reference: PFAMB; PB041022;
			Database reference: PFAMB; PB041022; Database reference: PFAMB; PB041498;
			Number of members: 271
			D500070
RNA_dep_RNApol2	!	RNA dependent RNA	Accession number: PF00978 Definition: RNA dependent RNA polymerase
		polymerase	- DD D-1 A
			Alignment method of seed: Clustalw
		1	Source of seed members: Pfam-B_13 (release 3.0)
			Gathering cutoffs: 8.5 0
			Trusted cutoffs: 8.50 0.20
			Noise cutoffs: 8 40 8 40
			HMM build command line: hmmbuild -f HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 93188140
			Reference Title: Roles of nonstructural polyproteins and
			cleavage products Reference Title: in regulating Sindbis virus RNA replication
			and Reference Title: transcription.
			Reference Author: Lemm JA, Rice CM;
			Reference Location: J Virol 1993;67:1916-1926.
			Reference Number: [2]
	1		Peterence Medline: 96323143
			Reference Title: Complete replication in vitro of tobacco
			mosaic virus RNA
			Reference Title: by a template-dependent, membrane-boun
			RNA polymerase.
			Reference Author: Osman TA, Buck KW;
			Reference Location: J Virol 1996;70:6227-6234.
			Reference Number: [3] Reference Medline: 94047331
			mana di alla and
			transcription require
	1		transcription require Reference Title: compatibility between the polymerase- and
			helicase-like
			Deference Title: viral RNA synthesis proteins.
			Reference Author: Dinant S, Janda M, Kroner PA, Ahlquist
			Reference Location: J Virol 1993;67:7181-7189.
			Reference Number: [4]
1	1		
	1		Reference Medline: 94094568 Reference Title: Evolution and taxonomy of positive-strand

\sim	\sim
ч	/
	9

Pfam	Prosite		992
riaii)	Prosite	Full Name	Description RNA viruses:
			Reference Title: implications of comparative analysis of
			amino acid Reference Title: sequences.
			Reference Little: sequences. Reference Author: Koonin EV, Dolja VV;
			Reference Location: Crit Rev Biochem Mol Biol 1993;28:375-430.
			Database Reference INTERPRO; IPR001788;
			Database reference: PFAMB; PB000096; Database reference: PFAMB; PB006751;
			Comment: This family may represent an RNA
			dependent RNA polymerase. Comment: The family contains the following proteins:
			Comment: 2A protein from bromoviruses
			Comment: putative RNA dependent RNA polymerase
			from tobamoviruses Comment: Non structural polyprotein from togaviruses
			Number of members: 125
RNA_pol	PDOC00410	Bacteriophage-type RNA	Many forms of RNA polymerase (EC 2.7.7.6) are known. Most
		polymerase family active site signatures	RNA polymerases are multimeric enzymes, but there is a family of single chain
			polymerases, which
			are evolutionary related, and which originate from bacteriophages or from
			mitochondria. The RNA polymerases that belong to this family
			are [1]:
			- Podoviridae bacteriophages T3, T7, and K11 polymerase.
			- Bacteriophage SP6 polymerase.
			 Vertebrate mitochondrial polymerase (gene POLRMT). Fungal mitochondrial polymerase (gene RPO41).
			- Polymerases encoded on mitochondrial linear DNA plasmids
			in various fungi
			and plants: Agaricus bitorquis pEM, Claviceps purpurea pClK1, Neurospora
			crassa Kalilo; Neurospora intermedia Maranhar and maize S-2).
			Two conserved aspartate and one lysine residue have been
			shown [2,3] to be part of the active site of T7 polymerase. We have used the
			regions around the
			first aspartate and around the lysine as signature patterns for this
			family of polymerases.
			Description of potters (a) and (a) mustile (a)
			Description of pattern(s) and/or profile(s)
			Consensus pattern P-[LIVM]-x(2)-D-[GA]-[ST]-[AC]-[SN]-[GA]-
			[LIVMFY]-Q [D is the active site residue] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern [LIVMF]-x-R-x(3)-K-x(2)-[LIVMF]-M-[PT]-x(2)-Y [K is the active site residue]
			Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update
			July 1999 / Text revised. References
			[1]
			McAllister W.T., Raskin C.A. Mol. Microbiol. 10:1-6(1993).
			[2]
			Maksimova T.G., Mustayev A.A., Zaychikov E.F., Lyakhov D.L.,
			Tunitskaya V.L., Akbarov A.K., Luchin S.V., Rechinsky V.O., Chernov B.K., Kochetkov S.N.
	1		

Pfam	Prosite	Full Name	Description
			[3]
			Sousa R., Chung Y.J., Rose J.P., Wang BC.
			Nature 364:593-599(1993).
RNA pol A	_	RNA polymerase alpha	Accession number: PF00623
rttvapoi_A		subunit	Definition: RNA polymerase alpha subunit
		ous a.m.	Author: Bateman A
			Alignment method of seed: HMM_built_from_alignment
			Source of seed members: Pfam-B_3 (release 2.1)
			Gathering cutoffs: 90
			Trusted cutoffs: 13.50 2.90
			Noise cutoffs: 8.50 8.50 HMM build command line: hmmbuild -f HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 97066998
			Reference Title: Structural modules of the large subunits of
			RNA polymerase.
			Reference Title: Introducing archaebacterial and chloroplast
			split sites in
			Reference Title: the beta and beta' subunits of Escherichia
			coli RNA Reference Title: polymerase.
			Reference Title: polymerase. Reference Author: Severinov K, Mustaev A, Kukarin A,
			Muzzin O, Bass I, Darst
			Reference Author: SA, Goldfarb A;
			Reference Location: J Biol Chem 1996;271:27969-27974.
			Database Reference INTERPRO; IPR000722;
			Database reference: PFAMB; PB003218;
			Comment: -!- RNA polymerases catalyse the DNA
			dependent polymerisation
			Comment: of RNA. Prokaryotes contain a single RNA polymerase
			Comment: compared to three in eukaryotes (not
			including mitochondrial.
			Comment: and chloroplast polymerases).
			Comment: -!- Members of this family include:
	İ		Comment: A subunit from eukaryotes
			Comment: gamma subunit from cyanobacteria
			Comment: beta' subunit from eubacteria
			Comment: A' subunit from archaebacteria
			Comment: B" from chloroplasts Number of members: 202
RNA pol A2		RNA polymerase	Accession number: PF01854
		A/beta'/A" subunit	Definition: RNA polymerase A/beta'/A" subunit
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_288 (release 4.2)
			Gathering cutoffs: -120 -120 Trusted cutoffs: -116.50 -116.50
			Noise cutoffs: -125.00 -125.00
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 88335550
			Reference Title: Relatedness of archaebacterial RNA
			polymerase core subunits
			Reference Title: to their eubacterial and eukaryotic
			equivalents. Reference Author: Berghofer B, Krockel L, Kortner C, Truss
			Reference Author: Berghofer B, Krockel L, Kortner C, Truss M, Schallenberg J,
			Reference Author: Klein A;
			Reference Location: Nucleic Acids Res 1988;16:8113-8128.
			Database Reference INTERPRO; IPR002879;
			Database reference: PFAMB; PB000546;
			Database reference: PFAMB; PB000846;
			Database reference: PFAMB; PB000984;
			Database reference: PFAMB; PB001168;
			Comment: RNA polymerases catalyse the DNA
			dependent polymerisation
			Comment: of RNA. Prokaryotes contain a single RNA
			polymerase Comment: compared to three in eukaryotes (not
			Compared to timee in editaryotes (not



Cheng Z.-F., Zuo Y., Li Z., Rudd K.E., Deutscher M.P.

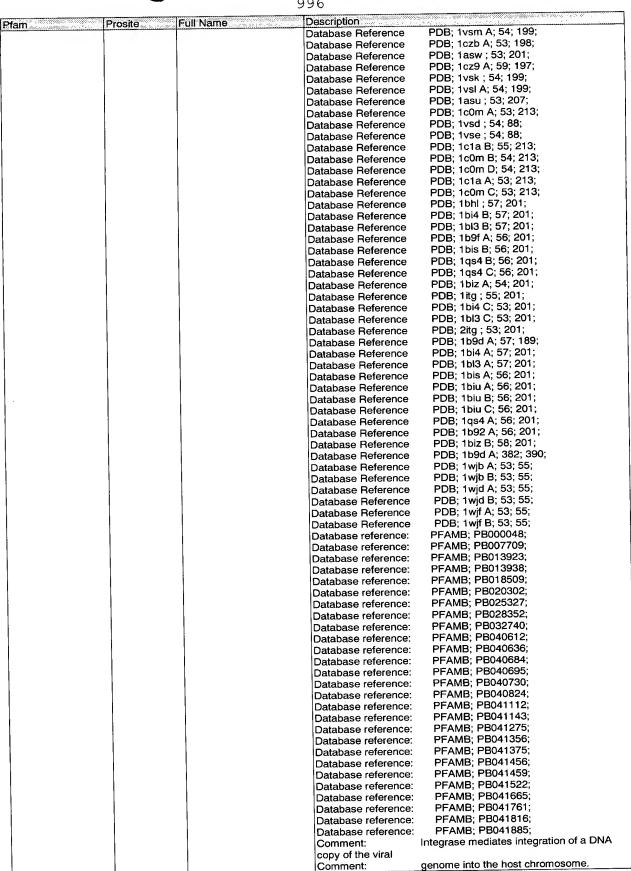
J. Biol. Chem. 273:14077-14080(1998).

[3]



			
Pfam	Prosite	Full Name	Description V. Cali T. Nakarawa M.
			Noguchi E., Hayashi N., Azuma Y., Seki T., Nakamura M.,
			Nakashima N., Yanagida M., He X., Mueller U., Sazer S., Nishimoto T.
			EMBO J. 15:5595-5605(1996).
			[4]
			Beuf L., Bedu S., Cami B., Joset F.
			Plant Mol. Biol. 27:779-788(1995).
			[5]
			Mian I.S.
			Nucleic Acids Res. 25:3187-3195(1997).
			` '
RRF		Ribosome recycling	Accession number: PF01765
		factor	Definition: Ribosome recycling factor
			Author: Bashton M, Bateman A Alignment method of seed: Clustalw
			Source of seed members: Pfam-B 949 (release 4.2)
			Gathering cutoffs: -35 -35
			Trusted cutoffs: -34.90 -34.90
			Noise cutoffs: -76.20 -76.20
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			Reference Medline: 94240115
			Reference Title: Ribosome recycling factor (ribosome
			releasing factor) is
			Reference Title: essential for bacterial growth.
			Reference Author: Janosi L, Shimizu I, Kaji A; Reference Location: Proc Natl Acad Sci U S A 1994;91:4249-
			Reference Location: Proc Natl Acad Sci U S A 1994;91:4249-4253.
			Database Reference INTERPRO; IPR002661;
			Comment: The ribosome recycling factor (RRF /
	,		ribosome release factor) dissociates
			Comment: the ribosome from the mRNA after
			termination of translation, and is
			Comment: essential bacterial growth [1]. Thus ribosomes are "recycled" and ready
			Comment: for another round of protein synthesis.
			Number of members: 27
	:		Number of members: 27
rve		Integrase core domain	Number of members: 27 Accession number: PF00665
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain Author: Bateman A
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases
rve		Integrase core domain	Number of members: 27 Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Title: [see comments] Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R,
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Title: [see comments] Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Reference Author: Davies DR;
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Title: [see comments] Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Reference Author: Davies DR; Reference Location: Science 1994;266:1981-1986.
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Title: [see comments] Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Reference Author: Davies DR; Reference Location: Science 1994;266:1981-1986. Database Reference: SCOP; 2itg; fa; [SCOP-USA][CATH-
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Title: [see comments] Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Reference Author: Davies DR; Reference Location: Science 1994;266:1981-1986. Database Reference: SCOP; 2itg; fa; [SCOP-USA][CATH-PDBSUM]
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Reference Author: Davies DR; Reference Author: Science 1994;266:1981-1986. Database Reference: SCOP; 2itg; fa; [SCOP-USA][CATH-PDBSUM] Database Reference
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Title: [see comments] Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Reference Author: Comments Science 1994;266:1981-1986. Database Reference: SCOP; 2itg; fa; [SCOP-USA][CATH-PDBSUM]
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Title: [see comments] Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Reference Author: Davies DR; Reference Author: Science 1994;266:1981-1986. Database Reference Databa
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Author: See comments] Reference Author: Davies DR; Reference Author: Davies DR; Reference Author: Science 1994;266:1981-1986. Database Reference
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Author: Integrase: similarity to other polynucleotidyl transferases Reference Author: Davies DR; Reference Author: Davies DR; Reference Author: Science 1994;266:1981-1986. Database Reference Data
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Reference Author: Davies DR; Reference Author: Science 1994;266:1981-1986. Database Reference
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Author: Davies DR; Reference Author: Davies DR; Reference Author: Davies DR; Reference Author: Science 1994;266:1981-1986. Database Reference Database Refere

	4	
	I	
	Ţ	
	ſ	
	Ī	
		,
L	-	
W		
		11111
		11111





		9	97
Pfam	Prosite	Full Name	Description
			Integrase is composed of Comment: three domains. The amino-terminal domain
			Comment: three domains. The amino-terminal domain is a zinc binding
			Comment: domain Integrase Zn. This domain is the
			central catalytic
			Comment: domain. The carboxyl terminal domain that
			is a non-specific
			Comment: DNA binding domain integrase.
			Comment: The catalytic domain acts as an
			endonuclease when two Comment: nucleotides are removed from the 3' ends of
			Comment: nucleotides are removed from the 3' ends of the blunt-ended
			Comment: viral DNA made by reverse transcription.
			This domain also
			Comment: catalyses the DNA strand transfer reaction
			of the 3' ends
			Comment: of the viral DNA to the 5' ends of the
			integration site [1].
-			Number of members: 1147
<u></u>		S4 domain	Accession number: PF01479
S4		UT UUIIIAIII	Definition: S4 domain
			Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Medline:99193178
			Gathering cutoffs: 17 17
			Trusted cutoffs: 17.20 17.20
			Noise cutoffs: 16.70 16.70
			HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 99193178
			Reference Title: Novel predicted RNA-binding domains
			associated with the
			Reference Title: translation machinery.
			Reference Author: Aravind L, Koonin EV;
			Reference Location: J Mol Evol 1999;48:291-302.
			Reference Number: [2]
			Reference Medline: 98372721 Reference Title: 98372721 The crystal structure of ribosomal protein
			S4 reveals a
			Reference Title: two-domain molecule with an extensive
			RNA-binding surface:
			Reference Title: one domain shows structural homology to
			the ETS DNA-binding
			Reference Title: motif.
			Reference Author: Davies C, Gerstner RB, Draper DE,
			Ramakrishnan V, White SW; Reference Location; EMBO J 1998;17:4545-4558.
	1		Database Reference: SCOP; 1c06; fa; [SCOP-USA][CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR002942;
			Database Reference PDB; 1c05 A; 51; 98;
			Database Reference PDB; 1c06 A; 51; 98;
			Database Reference PDB; 1dm9 A; 9; 55;
	1		Database Reference PDB; 1dm9 B; 9; 55; Database reference: PFAMB; PB001751;
			Database reference: PFAMB; PB001751; Database reference: PFAMB; PB041147;
			Database reference: PFAMB; PB041148;
			Comment: The S4 domain is a small domain consisting
			of 60-65 amino acid residues
			Comment: that was detected in the bacterial ribosomal
			protein S4, eukaryotic
			Comment: ribosomal S9, two families of pseudouridine
	1		synthases, a novel family
			Comment: of predicted RNA methylases, a yeast
	1		protein containing a pseudouridine Comment: synthetase and a deaminase domain,
			Comment: synthetase and a deaminase domain, bacterial tyrosyl-tRNA synthetases,
			Comment: and a number of uncharacterized, small
			proteins that may be involved in
	1	1	
		1	Comment: translation regulation [1]. The S4 domain
			Comment: translation regulation [1]. The S4 domain probably mediates binding to



			998
Pfam	Prosite	Full Name	Description
			Number of members: 256
SAA_proteins	PDOC00762	Serum amyloid A proteins signature	The serum amyloid A (SAA) proteins comprise a family of vertebrate proteins that associate predominantly with high density lipoproteins (HDL) [1,2]. The synthesis of certain members of the family is greatly increased (as much as a 1000 fold) in inflammation; thus making SAA a major acute phase reactant. While the major physiological function of SAA is unclear, prolonged elevation of plasma SAA levels, as in chronic inflammation, however, results in a pathological condition, called amyloidosis, which affects the liver, kidney and spleen and which is characterized by the highly insoluble accumulation of SAA in these tissues. SAA are proteins of about 110 amino acid residues. As a signature pattern, we selected the most highly conserved region, which is located in the central part of the sequence.
	-		Description of pattern(s) and/or profile(s) Consensus pattern A-R-G-N-Y-[ED]-A-x-[QKR]-R-G-x-G-G-x-W-A Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update June 1994 / First entry. References [1] Malle E., Steinmetz A., Raynes J.G. Atherosclerosis 102:131-146(1993). [2] Uhlar C.M., Burgess C.J., Sharp P.M., Whitehead A.S. Genomics 19:228-235(1994).
SAM		SAM domain (Sterile alpha motif)	Accession number: PF00536 Definition: SAM domain (Sterile alpha motif) Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1],[2] Gathering cutoffs: 11 0 Trusted cutoffs: 11.00 3.70 Noise cutoffs: 10.90 10.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 96100659 Reference Title: SAM: A novel motif in yeast sterile alpha and Drosophila Reference Title: polyhomeotic proteins Reference Author: Ponting CP; Reference Number: [2] Reference Medline: 97160498 SAM as a protein interaction domain involved in Reference Title: developmental regulation. Shultz J, Ponting CP, Hofmann K, Bork P; Prot Sci 1997;6:249-253. [3] Seference Medline: Prot Sci 1997;6:249-253. [3] Prot Sci 1997;6:249-253. [3] Prot Sci 1997;6:249-253. [3] Prot Sci 1997;6:249-253. [3] Prot Sci 1997;6:249-253. [3] Prot Sci 1997;6:249-253. [3] Prot Sci 1997;6:249-253. [3] Prot Sci 1997;6:249-253. [3] Prot Sci 1997;6:249-253. [3] Prot Sci 1997;6:249-253. [4] Prot Sci 1997;6:249-253. [5] Prot Sci 1997;6:249-253. [6] Prot Sci 1997;6:249-253. [7] Prot Sci 1997;6:249-253. [8] Prot Sci 1997;6:249-253. [9] Prot Sci 1997;6:249-253.



		3	99
Yam	Prosite	Full Name	Description P.S. I. I. B T. Siebei E.
			Reference Author: Stapleton D, Balan I, Pawson T, Sicheri F;
			Reference Location: Nat Struct Biol 1999;6:44-49.
			Database reference: SMART; SAM; Database Reference: SCOP; 1b0x; fa; [SCOP-USA][CATH-
			T
·			PDBSUM] Database Reference INTERPRO; IPR001660;
			Database Reference PDB; 1b0x A; 910; 973;
			Database Reference PDB; 1sgg; 7; 70;
			Database Reference PDB; 1b4f A; 7; 71;
			Database Reference PDB; 1b4f C; 7; 71;
			Database Reference PDB; 1b4f E; 7; 71;
			Database Reference PDB; 1b4f D; 7; 71;
			Database Reference PDB; 1b4f H; 7; 71;
		_	Database Reference PDB; 1b4f F; 7; 71;
			Database Reference PDB; 1b4f G; 7; 71;
			Database Reference PDB; 1b4f B; 7; 71; Database reference: PFAMB; PB008631;
		1	Database reference: PFAMB; PB008631; Database reference: PFAMB; PB040678;
			Database reference: PFAMB; PB041111;
			Database reference: PFAMB; PB041385;
			Comment: It has been suggested that SAM is an
			evolutionarily conserved protein
			Comment: binding domain that is involved in the
			regulation of numerous
			Comment: developmental processes in diverse
			eukaryotes.
			Comment: The SAM domain can potentially function as
			a protein interaction Comment: module through its ability to homo- and
			Comment: module through its ability to nomo- and
ļ			Comment: other SAM domains.
			Number of members: 110
			DE04500
SAM_decarbox		Adenosylmethionine	Accession number: PF01536
		decarboxylase	Definition: Adenosylmethionine decarboxylase Author: Bashton M, Bateman A
			Author: Bashton M, Bateman A Alignment method of seed: Clustalw
		1	Source of seed members: Pfam-B_600 (release 4.0)
			Gathering cutoffs: 11 11
			Trusted cutoffs: 17.90 17.90
			Noise cutoffs: 5.70 5.70
			HMM build command line: hmmbuild -f HMM SEED
			HMM build command line: hmmcalibrate seed 0 HMM
			Reference Number: [1]
			Reference Medline: 98098079
			Reference Title: Cloning, mapping and mutational analysis
			of the Reference Title: S-adenosylmethionine decarboxylase gene
			in Drosophila Reference Title: melanogaster.
			Reference Author: Larsson J, Rasmuson-Lestander A;
			Reference Location: Mol Gen Genet 1997;256:652-660.
			Database Reference: SCOP; 1jen; fa; [SCOP-USA][CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR001985;
		İ	Database Reference PDB; 1jen C; 69; 328;
			Database Reference PDB; 1jen A; 69; 329;
			Database Reference PDB; 1jen B; 4; 67;
			Database Reference PDB; 1jen D; 5; 66;
			Comment: This is a family of S-adenosylmethionine
			decarboxylase (SAMDC) proenzymes. Comment: In the biosynthesis of polyamines SAMDC
1			
			produces decarboxylated Comment: S-adenosylmethionine, which serves as the
	1		aminopropyl moiety necessary
l .			Comment: for spermidine and spermine biosynthesis
I	1		from putrescine [1]. The Pfam
		i e	
			Comment: alignment contains both the alpha and beta
			chains that are cleaved to
			chains that are cleaved to Comment: form the active enzyme.
			chains that are cleaved to
		0-4:	chains that are cleaved to Comment: form the active enzyme. Number of members: 34
SBF		Sodium Bile acid symporter family	chains that are cleaved to Comment: form the active enzyme.

Pfam	Prosite Full Name	Description
Jan - San Albania -	1 TOOKS 1 CHI I TOOMS	Author: Bashton M, Bateman A
		Alignment method of seed: Clustalw
		Source of seed members: Pfam-B_697 (release 4.2)
		Gathering cutoffs: -19 -19
		Trusted cutoffs: -12.50 -12.50
		Noise cutoffs: -26.40 -26.40
		HMM build command line: hmmbuild -F HMM SEED
		HMM build command line: hmmcalibrateseed 0 HMM
		Reference Number: [1]
		Reference Medline: 97377989 Reference Title: Isolation of three contiguous genes, ACR1,
		ACR2 and ACR3, Reference Title: involved in resistance to arsenic
		Reference Title: involved in resistance to arsenic compounds in the yeast
		Reference Title: Saccharomyces cerevisiae.
		Reference Author: Bobrowicz P, Wysocki R, Owsianik G,
		Goffeau A, Ulaszewski
		Reference Author: S;
		Reference Location: Yeast 1997;13:819-828.
		Reference Number: [2]
		Reference Medline: 92073340
		Reference Title: Functional expression cloning and
		characterization of the
		Reference Title: hepatocyte Na+/bile acid cotransport
		system.
		Reference Author: Hagenbuch B, Stieger B, Foguet M,
		Lubbert H, Meier PJ;
		Reference Location: Proc Natl Acad Sci U S A
		1991;88:10629-10633.
		Database Reference INTERPRO; IPR002657;
		Database reference: PFAMB; PB041594;
		Comment: This family consists of Na+/bile acid co-
		transporters.
		Comment: These transmembrane proteins function in
		the liver
		Comment: in the uptake of bile acids from portal blood
		plasma
		Comment: a process mediated by the co-transport of
		Na+ [2].
		Comment: Also in the family is ARC3 from S.
		cerevisiae Swiss:Q06598
		Comment: this is a putative transmembrane protein
وخي ١		involved in
		Comment: resistance to arsenic compounds [1].
		Number of members: 29
		BE01000
Sec7	Sec7 domain	Accession number: PF01369
		Definition: Sec7 domain
		Author: Bateman A
		Alignment method of seed: Clustalw_manual
		Source of seed members: Pfam-B_1629 (release 3.0)
		Gathering cutoffs: 25 25
		Trusted cutoffs: 101.50 101.50
		Noise cutoffs: 13.20 13.20
		HMM build command line: hmmbuild -f HMM SEED
		HMM build command line: hmmcalibrateseed 0 HMM
		Reference Number: [1] Reference Medline: 98169075
		exchange factor Reference Title: ARNO.
		Reference Title: ARNO. Reference Author: Cherfils J, Menetrey J, Mathieu M, Le
		Bras G, Robineau S,
		Reference Author: Beraud-Dufour S, Antonny B, Chardin F
		Reference Location: Nature 1998;392:101-105.
		Reference Number: [2]
		Reference Medline: 97100951
		Reference Title: A human exchange factor for ARF contain
		Sec7- and
		Reference Title: pleckstrin- homology domains.
		Reference Author: Chardin P, Paris S, Antonny B, Robinea
		S, Beraud-Dufour S,
	1	
	1	Reference Author: Jackson CL, Chabre M

Pfam	Prosite	Fulf Name	Description
			Satabase Reference: SCOP; 1pb/ (SCOP-USA][CATH-PDBSUM]
			Database Reference INTERPRO; IPR000904;
			Database Reference PDB; 1pbv; 58; 243;
			Database Reference PDB; 1bc9; 59; 244;
			Comment: The Sec7 domain is a guanine-nucleotide-
		İ	exchange-factor (GEF)
			Comment: for the arf family [2].
			Number of members: 32
Seedstore_2S		2S seed storage family	Accession number: PF01631
			Definition: 2S seed storage family
		·	Author: Bateman A Alignment method of seed: Clustalw
			Source of seed members: Pfam-B 1154 (release 4.1)
			Gathering cutoffs: 25 25
			Trusted cutoffs: 95.10 95.10
			Noise cutoffs: -0.20 10.10
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 97121264 Reference Title: 1H NMR assignment and global fold of
			napin Bnlb, a
			Reference Title: representative 2S albumin seed protein.
			Reference Author: Rico M, Bruix M, Gonzalez C, Monsalve
			RI, Rodriguez R;
			Reference Location: Biochemistry 1996;35:15672-15682.
			Database Reference: SCOP; 1pnb; fa; [SCOP-USA][CATH-
			PDBSUM] Database Reference INTERPRO: IPR000617:
			Database Reference INTERPRO; IPR000617; Database reference: PFAMB; PB029622;
			Comment: Members of this family are composed of two
			chains (both included in
			Comment: the alignment), these are co-translated and
			later cleaved. The two
			Comment: chains are disulphide linked together.
			Number of members: 27
SH2	PDOC50001	Src homology 2 (SH2)	The Src homology 2 (SH2) domain is a protein domain of about
		domain profile	100 amino-acid
			residues first identified as a conserved sequence region
			between the
			oncoproteins Src and Fps [1]. Similar sequences were later found in many other
			intracellular signal-transducing proteins [2]. SH2 domains
			function as regulatory modules of intracellular signalling cascades by
			interacting with
			high affinity to phosphotyrosine-containing target peptides in a
			sequence- specific and strictly phosphorylation-dependent manner [3,4,5,6].
			The SH2 domain has a conserved 3D structure consisting of two alpha helices
			and six to seven beta-strands. The core of the domain is
			formed by a
			continuous beta-meander composed of two connected beta-
			sheets [7].
	1		
			loroteins:
			- Many vertebrate, invertebrate and retroviral cytoplasmic (non-
			receptor)
	1		protein tyrosine kinases. In particular in the Src, Abl, Bkt, Csk
			and ZAP70
			families of kinases.
			- Mammalian phosphatidylinositol-specific phospholipase C gamma-1 and -2. Two
		I	Iganina i ana E. i iio
			copies of the SH2 domain are found in those proteins in
			copies of the SH2 domain are found in those proteins in between the



	1	002
Pfam Prosite	Full Name	Description
		subunit Some vertebrate and invertebrate protein-tyrosine phosphatases Mammalian Ras GTPase-activating protein (GAP) Adaptor proteins mediating binding of guanine nucleotide exchange factors to growth factor receptors: vertebrate GRB2, Caenorhabditis elegans sem-5 and Drosophila DRK Mammalian Vav oncoprotein, a guanine-nucleotide exchange factor of the CDC24 family Miscellanous proteins interacting with vertebrate receptor protein tyrosine kinases: oncoprotein Crk, mammalian cytoplasmic proteins Nck, Shc STAT proteins (signal transducers and activators of transcription) Chicken tensin Yeast transcriptional control protein SPT6. The profile developed to detect SH2 domains is based on a structural alignment consisting of 8 gap-free blocks and 7 linker regions totaling
		consisting of 8 gap-free blocks and 7 linker regions totaling 92 match positions. Description of pattern(s) and/or profile(s) Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT protein tyrosine kinases JAK1 and JAK2. Expert(s) to contact by email Zvelebil M. marketa@ludwig.ucl.ac.uk Last update November 1995 / First entry. References [1] Sadowski I., Stone J.C., Pawson T. Mol. Cell. Biol. 6:4396-4408(1986).
•		Russel R.B., Breed J., Barton G.J. FEBS Lett. 304:15-20(1992). [3] Marangere L.E.M., Pawson T. J. Cell Sci. Suppl. 18:97-104(1994). [4] Pawson T., Schlessinger J. Curr. Biol. 3:434-442(1993).
		[5] Mayer B.J., Baltimore D. Trends Cell. Biol. 3:8-13(1993). [6] Pawson T. Nature 373:573-580(1995). [7] Kuriyan J., Cowburn D. Curr. Opin. Struct. Biol. 3:828-837(1993).
Shikimate_DH	Shikimate / quinate 5- dehydrogenase	Accession number: PF01488 Definition: Shikimate / quinate 5-dehydrogenase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B 336 (release 4.0)



		1		
Pfam	Prosite	Full Name	Description	
			Gathering cutoffs:	-50 -50
	:		Trusted cutoffs:	-48.00 -48.00 -82.00 -82.00
				line: hmmbuild -F HMM SEED
				line: hmmcalibrateseed 0 HMM
			Reference Number:	[1]
			Reference Medline:	96048023
			Reference Title:	The molecular biology of multidomain
			proteins. Selected Reference Title:	examples.
			Reference Author:	Hawkins AR, Lamb HK;
			Reference Location:	Eur J Biochem 1995;232:7-18.
			Database Reference	INTERPRO; IPR002907;
			Comment:	This family contains both shikimate and
			quinate dehydrogena	
			Comment: conversion of	Shikimate 5-dehydrogenase catalyses the
			Comment:	shikimate to 5-dehydroshikimate. This
			reaction is part of	Similate to b donyardentimeter.
			Comment:	the shikimate pathway which is involved in
			the biosynthesis	
			Comment:	of aromatic amino acids.
			Comment:	Quinate 5-dehydrogenase catalyses the
			conversion of Comment:	quinate to 5-dehydroquinate. This reaction
			is part of	quillate to 5-deligoroquillate. This reaction
			Comment:	the quinate pathway where quinic acid is
			exploited as	
			Comment:	a source of carbon in prokaryotes and
			microbial	
			Comment:	eukaryotes. Both the shikimate and quinate pathways
			share two common	Both the shikimate and quinate paniways
			Comment:	pathway metabolites 3-dehydroquinate and
			dehydroshikimate.	
			Number of members:	58
			0: () [4]	
Sigma54_factors	PDOC00593	Sigma-54 factors family	promote	bacterial transcription initiation factors that
		signatures and profile		core RNA polymerase to specific initiation
			sites and are	
		E .		
				alter the specificity of promoter
			then released. They recognition. Most	
			then released. They recognition. Most bacteria express a m	alter the specificity of promoter ultiplicity of sigma factors. Two of these
			then released. They recognition. Most bacteria express a m factors, sigma-	ultiplicity of sigma factors. Two of these
		·	then released. They recognition. Most bacteria express a m factors, sigma- 70 (gene rpoD), ger	
			then released. They recognition. Most bacteria express a m factors, sigma- 70 (gene rpoD), ger factor, and	ultiplicity of sigma factors. Two of these
		·	then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpoN variety of	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma
		·	then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpoN variety of genes. The other stactors, are	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpoN variety of genes. The other stactors, are	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma
		·	then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequi grouped into two	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma N or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes.
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpoD) variety of genes. The other sfactors, are required for the trans. With regard to sequing grouped into two classes: the sigma-5	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes.
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpoN variety of genes. The other sfactors, are required for the trans. With regard to sequing grouped into two classes: the sigma-5 family has many	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be 34 and sigma-70 families. The sigma-70
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpoN variety of genes. The other sfactors, are required for the trans. With regard to sequingrouped into two classes: the sigma-5 family has many different sigma factors.	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma N or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes.
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequi grouped into two classes: the sigma-5 family has many different sigma factor The sigma-54	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. ence similarity, sigma factors can be 34 and sigma-70 families. The sigma-70 res (see the relevant entry <pdoc00592>).</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequi grouped into two classes: the sigma-5 family has many different sigma factor The sigma-54	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be 34 and sigma-70 families. The sigma-70
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequipart grouped into two classes: the sigma-5 family has many different sigma factor the sigma-54 family consists exclifor the	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. ence similarity, sigma factors can be 34 and sigma-70 families. The sigma-70 res (see the relevant entry <pdoc00592>).</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpoD) variety of genes. The other s factors, are required for the trans. With regard to sequing grouped into two classes: the sigma-5 family has many different sigma factor The sigma-54 family consists excliptor the transcription of promisensus.	nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be and sigma-70 families. The sigma-70 are (see the relevant entry <pdoc00592>). Susively of sigma-54 factor [2,3] required noters that have a characteristic -24 and -12</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpoD) variety of genes. The other s factors, are required for the trans. With regard to sequing grouped into two classes: the sigma-54 family has many different sigma factor. The sigma-54 family consists excliptor the transcription of promisons consensus recognition element.	ultiplicity of sigma factors. Two of these nerally known as the major or primary sigma of or ntrA direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be a factor and a sigma-70 families. The sigma-70 are (see the relevant entry <pdoc00592>). Subsively of sigma-54 factor [2,3] required</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequi grouped into two classes: the sigma-54 family has many different sigma factor the sigma-54 family consists excl for the transcription of prom consensus recognition element sequences	nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be seen the relevant entry <pdoc00592>). The sigma-70 is sigma-54 factor [2,3] required noters that have a characteristic -24 and -12 but which are devoid of the typical -10,-35</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequi grouped into two classes: the sigma-5 family has many different sigma factor the sigma-54 family consists excl for the transcription of prom consensus recognition element sequences recognized by the	nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be and sigma-70 families. The sigma-70 are (see the relevant entry <pdoc00592>). Susively of sigma-54 factor [2,3] required noters that have a characteristic -24 and -12</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequing grouped into two classes: the sigma-54 family has many different sigma factor the sigma-54 family consists excl for the transcription of prom consensus recognition element sequences recognized by the is also	nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be seen a sigma-70 families. The sigma-70 ars (see the relevant entry <pdoc00592>). Lusively of sigma-54 factor [2,3] required the process that have a characteristic -24 and -12 and the process that have a characteristic -24 and -12 and the process that have a characteristic -24 and -13 and the process that have a characteristic -24 and -15 and the</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequing grouped into two classes: the sigma-54 family has many different sigma factor the sigma-54 family consists excl for the transcription of prom consensus recognition element sequences recognized by the is also	nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be seen the relevant entry <pdoc00592>). The sigma-70 is sigma-54 factor [2,3] required noters that have a characteristic -24 and -12 but which are devoid of the typical -10,-35</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpoD) variety of genes. The other sfactors, are required for the trans. With regard to sequing grouped into two classes: the sigma-54 family has many different sigma factor. The sigma-54 family consists exclifor the transcription of promisensus recognition element sequences recognized by the sigma-50 characterized by its regulatory	nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be seen a sigma-70 families. The sigma-70 ars (see the relevant entry <pdoc00592>). Lusively of sigma-54 factor [2,3] required the process of the sigma-12 and 12 are that have a characteristic -24 and -12 and which are devoid of the typical -10,-35 major sigma factors. The sigma-54 factor</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequity grouped into two classes: the sigma-54 family has many different sigma factor the sigma-54 family consists excl for the transcription of promictorsensus recognition element sequences recognized by the is also characterized by its regulatory proteins that bind to	nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be seen a sigma-70 families. The sigma-70 ars (see the relevant entry <pdoc00592>). Housively of sigma-54 factor [2,3] required to the table and the table and the table and the typical -10,-35 major sigma factors. The sigma-54 factor interaction with ATP-dependent positive</pdoc00592>
			then released. They recognition. Most bacteria express a m factors, sigma-70 (gene rpoD), ger factor, and sigma-54 (gene rpot variety of genes. The other s factors, are required for the trans. With regard to sequity grouped into two classes: the sigma-54 family has many different sigma factor The sigma-54 family consists excl for the transcription of prom consensus recognition element sequences recognized by the is also characterized by its regulatory proteins that bind to	nerally known as the major or primary sigma of or ntrA) direct the transcription of a wide sigma factors, known as alternative sigma cription of specific subsets of genes. Hence similarity, sigma factors can be seen the relevant entry <pdoc00592>). The sigma-70 families. The sigma-70 are (see the relevant entry <pdoc00592>). The sigma-54 factor [2,3] required to the sigma-54 factor sigma-54 factor sigma-54 factor. The sigma-54 factor interaction with ATP-dependent positive upstream activating sequences.</pdoc00592></pdoc00592>



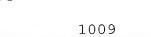
Pfam	Prosite	Full Name	Description
			about 50
			residues that contains a potential leucine zipper motif A region of variable length which is not well conserved A well conserved C-terminal region of about 350 residues that
			contains a second potential leucine zipper, a potential DNA-binding 'helix-turn-helix'
:			motif and a perfectly conserved octapeptide whose function is not known.
			We developed two signature patterns for this family of sigma factors. The
			first starts two residues before the N-terminal extremity of the helix-turn-
			helix region and ends two residues before its C-terminal extremity. The second
			is the conserved octapeptide. A profile has also been designed that covers the whole C-terminal region.
			Description of pattern(s) and/or profile(s)
			Consensus pattern P-[LIVM]-x-[LIVM]-x(2)-[LIVM]-A-x(2)- [LIVMFT]-x(2)-[HS]-x- S-T-[LIVM]-S-R Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern R-R-T-[IV]-[ATN]-K-Y-R Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update
			July 1999 / Patterns and text revised. References [1]
:			Helmann J.D., Chamberlin M.J. Annu. Rev. Biochem. 57:839-872(1988).
			[2] Thoeny B., Hennecke H. FEMS Microbiol. Rev. 5:341-358(1989).
			[3] Merrick M.J. Mol. Microbiol. 10:903-909(1993).
SLH	PDOC00823	S-layer homology domain signature	S-layers are paracrystalline mono-layered assemblies of (glyco)proteins which coat the surface of bacteria [1]. Several S-layer proteins and some other cell wall proteins contain one or more copies of a domain of about 50-
			60 residues, which has been called SLH (for S-layer homology) [2]. There is strong evidence that this domain serves as an anchor to the peptidoglycan [3]. The SLH domain
:			has been found in:
			 S-layer glycoprotein of Acetogenium kivui (3 copies). S-layer 125 Kd protein of Bacillus sphaericus (3 copies). S-layer protein of Bacillus anthracis (3 copies). S-layer protein of Bacillus licheniformis (3 copies).

fam Prosite Full Name Description - S-layer protein (HWP) from Bacillus brevis strain HPD31 (copies). - Middle cell wall protein (MWP) from Bacillus brevis strain copies). - S-layer protein (p100) of Thermus thermophilus (1 copy). - Outer membrane protein Omp-alpha from Thermotoga ma (1 copy). - Cellulosome anchoring protein (gene ancA), outer layer pr B (OlpB) and a further potential cell surface glycoprotein from Clostridiun thermocellum (3 copies; the first copy is missing its N-terminal third whice appended to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacter thermosulfurogenes (3 copies). - Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Exoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3 of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). - SLH domains are found at the N- or C-termini of mature pro They occur in	7 (3 rritima ottein n n is
copies). - Middle cell wall protein (MWP) from Bacillus brevis strain 4 copies). - S-layer protein (p100) of Thermus thermophilus (1 copy). - Outer membrane protein Omp-alpha from Thermotoga ma (1 copy). - Cellulosome anchoring protein (gene ancA), outer layer pr B (OlpB) and a further potential cell surface glycoprotein from Clostridiun thermocellum (3 copies; the first copy is missing its N-terminal third whice appended to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies). - Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Endoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical protein sfrom Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account).	7 (3 rritima ottein n n is
copies). - S-layer protein (p100) of Thermus thermophilus (1 copy). - Outer membrane protein Omp-alpha from Thermotoga ma (1 copy). - Cellulosome anchoring protein (gene ancA), outer layer pr B (OlpB) and a further potential cell surface glycoprotein from Clostridiun thermocellum (3 copies; the first copy is missing its N-terminal third whice appended to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies) - Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Exoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacillus circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account).	ritima otein n n is r
- Outer membrane protein Omp-alpha from Thermotoga ma (1 copy). - Cellulosome anchoring protein (gene ancA), outer layer properly a further potential cell surface glycoprotein from Clostridius thermocellum (3 copies; the first copy is missing its N-terminal third whice appended to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies) - Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Exoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature profiles are found at the N- or C-termini of mature profiles are found at the N- or C-termini of mature profiles.	otein n n is r
(1 copy). Cellulosome anchoring protein (gene ancA), outer layer promoted a further potential cell surface glycoprotein from Clostridium thermocellum (3 copies; the first copy is missing its N-terminal third whice appended to the end of the third copy; may have arisen by circular permutation). Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies) Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies). Endoglucanase from Bacillus strain KSM-635 (3 copies). Exoglucanase (gene xynA) from Clostridium thermocellum copies). Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). Two hypothetical proteins from Synechocystis strain PCC (1 copy each). A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature profiles are found at the N- or C-termini of mature profiles.	otein n n is r
B (OlpB) and a further potential cell surface glycoprotein from Clostridius thermocellum (3 copies; the first copy is missing its N-terminal third whice appended to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies) - Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies) Endoglucanase from Bacillus strain KSM-635 (3 copies) Exoglucanase (gene xynX) from Clostridium thermocellum copies) Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account) Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account) Two hypothetical proteins from Synechocystis strain PCC (1 copy each) A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro They occur in	n is r o1 (3
a further potential cell surface glycoprotein from Clostridius thermocellum (3 copies; the first copy is missing its N-terminal third whice appended to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies) - Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Exoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - Two hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account).	n is r o1 (3
thermocellum (3 copies; the first copy is missing its N-terminal third whic appended to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies) - Amylopullulanase (gene aapT) from Bacillus strain XAL-6(copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Exoglucanase (gene xynX) from Clostridium thermocellum copies). - Xytanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro They occur in	n is r o1 (3
appended to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies) - Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Exoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro They occur in	r o1 (3
to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies) - Amylopullulanase (gene aapT) from Bacillus strain XAL-66 copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Endoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature prothey occur in	1 (3
- Amylopullulanase (gene amyB) from Thermoanaerobacte thermosulfurogenes (3 copies) - Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies) Endoglucanase from Bacillus strain KSM-635 (3 copies) Exoglucanase (gene xynX) from Clostridium thermocellum copies) Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account) Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account) Two hypothetical proteins from Synechocystis strain PCC (1 copy each) A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature proteiney occur in	1 (3
thermosulfurogenes (3 copies) Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies). Endoglucanase from Bacillus strain KSM-635 (3 copies). Exoglucanase (gene xynX) from Clostridium thermocellum copies). Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). Two hypothetical proteins from Synechocystis strain PCC (1 copy each). A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro	1 (3
- Amylopullulanase (gene aapT) from Bacillus strain XAL-60 copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Exoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature protein years are found at the N- or C-termini of mature protein years.	
copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Exoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature protein years.	
- Exoglucanase (gene xynX) from Clostridium thermocellum copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature proteins of the control of the contr	(3
copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature protein years are found at the N- or C-termini of mature proteins from the control of the country in the control of the control of the copy is account to the copy in the copy in the copy in the copy in the copy is account to the copy in the cop	(O
saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro They occur in	
copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature protein the procur in	
- Protein involved in butirosin production (ButB) from Bacill circulans (2 incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature protein the procur in the sequence of the sequence similarity to amylopullulanases.	
incomplete copies; 3 copies if three frameshifts are taken account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature protein the control of the control	1S
account). - Two hypothetical proteins from Synechocystis strain PCC (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 c) 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro They occur in	nto
(1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro They occur in	
- A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro They occur in	5803
3 ^r of amylase gene from Bacillus circulans (fragment of 1 of 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro	
3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature pro They occur in	onv.
SLH domains are found at the N- or C-termini of mature pro	.ору,
They occur in	
They occur in	eins.
the state of the s	
single copy followed by a predicted coiled coil domain, or in contiguous	unee
copies. Structurally, the SLH domain is predicted to contain	two
alpha-helices flanking a beta strand. The SLH sequences are fairly diverg	ent
with an average	
identity of about 25%. It is however possible to build a seque	ence
starts at the second position of the domain and that spans 3	/4 of
its length.	
Description of pattern(s) and/or profile(s)	
Consensus pattern [LVFYT]-x-[DA]-x(2,5)-[DNGSATPHY]- [FYWPDA]-x(4)-[LIV]-x(2)- [GTALV]-x(4,6)-[LIVFYC]-x(2)-G	
[PGSTA]-x(2,3)-[MFYA]-x-[PGAV]-x(3,10)-[LIVMA]-[STKR]	-[RY]-
x-[EQ]-x-[STALIVM]	
Sequences known to belong to this class detected by the p	WC(1)
Other sequence(s) detected in SWISS-PROT NONE.	
Expert(s) to contact by email Lupas A.N. lupas@vms.biochem.mpg.de	
Last update November 1997 / Pattern and text revised.	
References	
[1] Beveridge T.J.	
Curr. Opin. Struct. Biol. 4:204-212(1994).	
IC 3	
[2] Lupas A., Engelhardt H., Peters J., Santarius U., Volker S.	

			Description
Pfam	Prosite	Full Name	Baumeister W.
			J. Bacteriol. 176:1224-1233(1994).
			,
			[3]
			Lemaire M., Ohayon H., Gounon P., Fujino T., Beguin P. J. Bacteriol. 177:2451-2459(1995).
			J. Bacteriol. 177.2451-2453(1555).
Smr		Smr domain	Accession number: PF01713
			Definition: Smr domain
			Author: Bateman A Alignment method of seed: Clustalw
			Source of seed members: [1]
			Gathering cutoffs: 0 0
			Trusted cutoffs: 1.40 1.40
			Noise cutoffs: -7.90 -7.90 HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 10431172 Reference Title: Smr: a bacterial and eukaryotic homologue
			Reference Title: Smr: a bacterial and eukaryotic nomologue of the C-terminal
			Reference Title: region of the MutS2 family.
			Reference Author: Moreira D, Philippe H;
			Reference Location: Trends Biochem Sci 1999;24:298-300. Database Reference INTERPRO; IPR002625;
			Comment: This family includes the Smr (Small MutS
i			Related) proteins.
			Comment: and the C-terminal region of the MutS2
İ			protein. It has been Comment: suggested that this domain interacts with
			the MutS1
			Comment: Swiss:P23909 protein in the case of Smr
			proteins and with Comment: the N-terminal MutS related region of MutS2
			Swiss:P94545 [1].
			Number of members: 14
		14400	A number of transcription factors contain a conserved domain of
SRF-TF	PDOC00302	MADS-box domain signature and profile	56 amino-acid
		Signature and prome	residues, sometimes known as the MADS-box domain [E1]. They
			are listed below:
			- Serum response factor (SRF) [1], a mammalian transcription
•			factor that
			binds to the Serum Response Element (SRE). This is a short
			sequence of dyad symmetry located 300 bp to the 5' end of the transcription
			initiation site
			of genes such as c-fos.
			- Mammalian myocyte-specific enhancer factors 2A to 2D (MEF2A to MEF2D).
			These proteins are transcription factor which binds specifically
			to the
			MEF2 element present in the regulatory regions of many muscle-specific
			genes.
			- Drosophila myocyte-specific enhancer factor 2 (MEF2).
			- Yeast GRM/PRTF protein (gene MCM1) [2], a transcriptional
			regulator of mating-type-specific genes.
			- Yeast arginine metabolism regulation protein I (gene ARGR1 or
			ARG80).
			- Yeast transcription factor RLM1 Yeast transcription factor SMP1.
			- Arabidopsis thaliana agamous protein (AG) [3], a probable
			transcription
			factor involved in regulating genes that determines stamen
			and carpel development in wild-type flowers. Mutations in the AG gene
			result in the
			replacement of the stamens by petals and the carpels by a new
			flower Arabidopsis thaliana homeotic proteins Apetala1 (AP1),
			Apetala3 (AP3) and
		_1	li ibaniana li ii al'ania

		Τ,	007
Pfam	Prosite	Full Name	Description Pistillata (PI) which act locally to specify the identity of the floral
			meristem and to determine sepal and petal development [4] Antirrhinum majus and tobacco homeotic protein deficiens (DEFA) and globosa
			(GLO) [5]. Both proteins are transcription factors involved in the genetic control of flower development. Mutations in DEFA or GLO
			cause the transformation of petals into sepals and of stamina into carpels.
			Arabidopsis thaliana putative transcription factors AGL1 to AGL6 [6]. Antirrhinum majus morphogenetic protein DEF H33 (squamosa).
			In SRF, the conserved domain has been shown [1] to be involved in DNA-binding and dimerization. We have derived a pattern that spans the
			complete length of the domain. The profile also spans the length of the MADS-box.
			Description of pattern(s) and/or profile(s)
			Consensus pattern R-x-[RK]-x(5)-I-x-[DNGSK]-x(3)-[KR]-x(2)-T-[FY]-x-[RK](3)- x(2)-[LIVM]-x-K(2)-A-x-E-[LIVM]-[STA]-x-L-x(4)-[LIVM]-x- [LIVM](3)-x(6)-[LIVMF]-x(2)-[FY] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Sequences known to belong to this class detected by the profile ALL.
			Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so.
			Last update July 1999 / Pattern and text revised. References [1]
			Norman C., Runswick M., Pollock R., Treisman R. Cell 55:989-1003(1988).
			[2] Passmore S., Maine G.T., Elble R., Christ C., Tye BK. J. Mol. Biol. 204:593-606(1988).
			[3] Yanofsky M., Ma H., Bowman J., Drews G., Feldmann K.A., Meyerowitz E.M. Nature 346:35-39(1990).
			[4] Goto K., Meyerowitz E.M. Genes Dev. 8:1548-1560(1994).
			[5] Troebner W., Ramirez L., Motte P., Hue I., Huijser P., Loennig W. E., Saedler H., Sommer H., Schwartz-Sommer Z. EMBO J. 11:4693-4704(1992).
			[6] Ma H., Yanofsky M.F., Meyerowitz E.M. Genes Dev. 5:484-495(1991).
			[E1] http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C0014
SRP19		SRP19 protein	Accession number: PF01922 Definition: SRP19 protein Author: Enright A, Ouzounis C, Bateman A

(A)	1		
Pfam	Prosite	Full Name	Description
			Alignment method of seed: Clustalw
			Source of seed members: Enright A
			Gathering cutoffs: 25 25
			Trusted cutoffs: 31.20 31.20
	1		Noise cutoffs: -28.50 -28.50
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 89041541
			Reference Title: Isolation and characterization of a cDNA
			clone encoding the
			Reference Title: 19 kDa protein of signal recognition particle (SRP):
			Reference Title: expression and binding to 7SL RNA. Reference Author: Lingelbach K, Zwieb C, Webb JR,
			Marshallsay C, Hoben PJ,
			Reference Author: Walter P, Dobberstein B;
			Reference Location: Nucleic Acids Res 1988;16:9431-9442.
	1		Reference Number: [2]
			Reference Medline: 92220168
			Reference Title: SEC65 gene product is a subunit of the
			yeast signal
			Reference Title: recognition particle required for its integrity.
			Reference Author: Hann BC, Stirling CJ, Walter P;
			Reference Location: Nature 1992;356:532-533.
			Reference Number: [3]
			Reference Medline: 92220169
	1		Reference Title: The S. cerevisiae SEC65 gene encodes a
			component of yeast
			Reference Title: signal recognition particle with homology to
			human SRP19.
			Reference Author: Stirling CJ, Hewitt EW;
			Reference Location: Nature 1992;356:534-537.
			Database Reference INTERPRO; IPR002778;
			Comment: The signal recognition particle (SRP) binds
	1		to the signal peptide of
			Comment: proteins as they are being translated. The
			binding of the SRP halts
			l = "
			· ·
			transported to the endoplasmic Comment: reticulum's cytoplasmic surface. The SRP
			Comment: reticulum's cytoplasmic surface. The SRP then aids translocation of
			1
			Comment: the protein through the ER membrane. The
			SRP is a ribonucleoprotein
			Comment: that is composed of a small RNA and
			several proteins. One of these
			Comment: proteins is the SRP19 protein [1] (Sec65 in
			yeast [2,3]).
			Number of members: 13
CCD	DDOOCCCC	Cinala atana di Li	The Forbacietic and single strend bindles and 1997
SSB	PDOC00602	Single-strand binding	The Escherichia coli single-strand binding protein [1] (gene ssb),
		protein family signatures	also known
			as the helix-destabilizing protein, is a protein of 177 amino
			acids. It
			binds tightly, as a homotetramer, to single-stranded DNA (ss-
			DNA) and plays an
	1		important role in DNA replication, recombination and repair.
			Closely related variants of SSB are encoded in the genome of a variety of
			large self-transmissible plasmids. SSB has also been
			characterized in bacteria
			such as Proteus mirabilis or Serratia marcescens.
			Eukaryotic mitochondrial proteins that bind ss-DNA and are
			probably involved
			in mitochondrial DNA replication are structurally and evolutionary related to
			prokaryotic SSB. Proteins currently known to belong to this
	1		subfamily are
	1		listed below [2].
		i	
		1	- Mammalian protein Mt-SSB (P16).
			- Xenopus Mt-SSBs and Mt-SSBr.



			009
Pfam	Prosite	Full Name	Description Description
			- Drosophila MtSSB. - Yeast protein RIM1.
		•	- reast protein rater.
			We have developed two signature patterns for these proteins.
			The first is a
			conserved region in the N-terminal section of the SSB's. The second is a
			centrally located region which, in Escherichia coli SSB, is
			known to be
			involved in the binding of DNA.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIVMF]-[NST]-[KRHST]-[LIVM]-x-[LIVMF](2)-
			G-[NHRK]- [LIVMA]-[GST]-x-[DENT]
			Sequences known to belong to this class detected by the pattern
			ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern T-x-W-[HY]-[RNS]-[LIVM]-x-[LIVMF]-[FY]-
			[NGKR]
			Sequences known to belong to this class detected by the pattern
			A majority. Other sequence(s) detected in SWISS-PROT NONE.
			Last update
			December 1999 / Patterns and text revised.
			References
			[1] Meyer R.R., Laine P.S.
			Microbiol. Rev. 54:342-380(1990).
			,
			[2]
			Stroumbakis N.D., Li Z., Tolias P.P. Gene 143:171-177(1994).
			Gene 143.771-177(1334).
START		START domain	Accession number: PF01852
			Definition: START domain
			Author: SMART Alignment method of seed: Manual
			Source of seed members: Alignment kindly provided by SMART
			Gathering cutoffs: 25 25
	-		Trusted cutoffs: 106.20 106.20
			HMM build command line: hmmbuild HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
	1		Reference Medline: 99257451 Reference Title: START: a lipid-binding domain in StAR,
			HD-ZIP and
			Reference Title: signalling proteins.
			Reference Author: Ponting CP, Aravind L; Reference Location: Trends Biochem Sci 1999;24:130-132.
			Database reference: SMART; START;
			Database Reference INTERPRO; IPR002913;
1			INC. I are also a mark many and
			Number of members: 41
Observation of the control of the co		Storol docaturaco	
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1)
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 12.90 12.90 Noise cutoffs: -44.50 -44.50
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 12.90 12.90 Noise cutoffs: -44.50 -44.50 HMM build command line: hmmbuild -F HMM SEED
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 12.90 12.90 Noise cutoffs: -44.50 -44.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 12.90 12.90 Noise cutoffs: -44.50 -44.50 HMM build command line: hmmbuild -F HMM SEED
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 12.90 12.90 Noise cutoffs: -44.50 -44.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 91323727 Reference Title: Cloning, disruption and sequence of the
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 12.90 12.90 Noise cutoffs: -44.50 -44.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 91323727 Reference Title: Cloning, disruption and sequence of the gene encoding yeast
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 12.90 12.90 Noise cutoffs: -44.50 -44.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 91323727 Reference Title: Cloning, disruption and sequence of the

Pfam	Prosite	Full Name	Description
<u></u>			Guynn CJ, Barbuch
			Reference Author: RJ, Ulbright CE, Bard M;
			Reference Location: Gene 1991;102:39-44.
			Reference Number: [2]
			Reference Medline: 96133902
			Reference Title: Cloning and characterization of ERG25, the
			Saccharomyces
			Reference Title: cerevisiae gene encoding C-4 sterol methyl
			oxidase. Reference Author: Bard M, Bruner DA, Pierson CA, Lees
		1	,
			ND, Biermann B, Frye L, Reference Author: Koegel C, Barbuch R;
			Reference Author: Koegel C, Barbuch H; Reference Location: Proc Natl Acad Sci U S A 1996;93:186-
			190.
			Reference Number: [3]
			Reference Medline: 96351930
			Reference Title: Molecular characterization of the CER1
			gene of arabidopsis
			Reference Title: involved in epicuticular wax biosynthesis
			and pollen
			Reference Title: fertility.
			Reference Author: Aarts MG, Keijzer CJ, Stiekema WJ,
			Pereira A;
			Reference Location: Plant Cell 1995;7:2115-2127.
			Database Reference INTERPRO; IPR001541;
			Database reference: PFAMB; PB041851;
			Comment: This family includes C-5 sterol desaturase
			and C-4 sterol methyl
			Comment: oxidase. Members of this family are
			involved in cholesterol biosynthesis
		i	Comment: and biosynthesis a plant cuticular wax.
			These enzymes contain many
			Comment: conserved histidine residues. Members of
			this family are integral
			Comment: mebrane proteins.
			Number of members: 34
			To a 1 a harden a vorious sulfato
Sulfatase	PDOC00117	Sulfatases signatures	Sulfatases (EC 3.1.6) are enzymes that hydrolyze various sulfate
		1	esters. The sequence of different types of sulfatases are available. These
			enzymes are:
			- Arylsulfatase A (EC 3.1.6.8) (ASA), a lysosomal enzyme which
			hydrolyzes
			cerebroside sulfate.
			- Arylsulfatase B (EC 3.1.6.12) (ASB), a lysosomal enzyme
			which hydrolyzes
			the sulfate ester group from N-acetylgalactosamine 4-sulfate
	1		, , , , , , , , , , , , , , , , , , ,
			residues of
			dermatan sulfate.
			7-5-1-1-1
			dermatan sulfate Arylsulfatase C (ASD) Arylsulfatase E (ASE).
			dermatan sulfate. - Arylsulfatase C (ASD).
			dermatan sulfate Arylsulfatase C (ASD) Arylsulfatase E (ASE) Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound
			dermatan sulfate Arylsulfatase C (ASD) Arylsulfatase E (ASE) Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a
			dermatan sulfate Arylsulfatase C (ASD) Arylsulfatase E (ASE) Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates.
			dermatan sulfate Arylsulfatase C (ASD) Arylsulfatase E (ASE) Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a
			dermatan sulfate Arylsulfatase C (ASD) Arylsulfatase E (ASE) Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme
			dermatan sulfate Arylsulfatase C (ASD) Arylsulfatase E (ASE) Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic
			dermatan sulfate Arylsulfatase C (ASD) Arylsulfatase E (ASE) Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate.
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate.
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan sulfate.
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan sulfate. - Choline sulfatase (EC 3.1.6.6) (gene betC), a bacterial
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan sulfate. - Choline sulfatase (EC 3.1.6.6) (gene betC), a bacterial enzyme that
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan sulfate. - Choline sulfatase (EC 3.1.6.6) (gene betC), a bacterial enzyme that converts choline-O-sulfate to choline.
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan sulfate. - Choline sulfatase (EC 3.1.6.6) (gene betC), a bacterial enzyme that converts choline-O-sulfate to choline. - Glucosamine-6-sulfatase (EC 3.1.6.14) (G6S), a lysosomal
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan sulfate. - Choline sulfatase (EC 3.1.6.6) (gene betC), a bacterial enzyme that converts choline-O-sulfate to choline. - Glucosamine-6-sulfatase (EC 3.1.6.14) (G6S), a lysosomal enzyme that
			dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6.4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan sulfate. - Choline sulfatase (EC 3.1.6.6) (gene betC), a bacterial enzyme that converts choline-O-sulfate to choline. - Glucosamine-6-sulfatase (EC 3.1.6.14) (G6S), a lysosomal



			1011
Pfam	Prosite	Full Name	Description
			and keratan sulfate. - N-sulphoglucosamine sulphohydrolase (EC 3.10.1.1) (sulphamidase), the lysosomal enzyme that catalyzes the hydrolysis of N-sulfo-d-
			glucosamine into glucosamine and sulfate.
			- Sea urchin embryo arylsulfatase (EC 3.1.6.1) Green alga arylsulfatase (EC 3.1.6.1), an enzyme which plays an important role in the mineralization of sulfates.
			- Arylsulfatase (EC 3.1.6.1) from Escherichia coli (gene asIA), Klebsiella
			aerogenes (gene atsA) and Pseudomonas aeruginosa (gene atsA). - Escherichia coli hypothetical protein yidJ.
			It has been shown that all these sulfatases are structurally related [1,2,3].
			As signature patterns for that family of enzymes we have selected the two best conserved regions. Both regions are located in the N-terminal
	:		section of these enzymes. The first region contains a conserved arginine which
			could be implicated in the catalytic mechanism; it is located four residues after a
			position that, in eukaryotic sulfatases, is a conserved cysteine which has been shown [4] to be modified to 2-amino-3-oxopropionic acid. In
			prokaryotes, this cysteine is replaced by a serine.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [SAP]-[LIVMST]-[CS]-[STAC]-P-[STA]-R-x(2)-[LIVMFW](2)- [TAR]-G [R is a putative active site residue] Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE.
	i		Consensus pattern G-[YV]-x-[ST]-x(2)-[IVAS]-G-K-x(0,1)- [FYWMK]-[HL]
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update December 1999 / Patterns and text revised. References
			[1] Peters C., Schmidt B., Rommerskirch W., Rupp K., Zuehlsdorf M., Vingron M., Meyer H.E., Pohlmann R., von Figura K. J. Biol. Chem. 265:3374-3381(1990).
			[2] Wilson P.J., Morris C.P., Anson D.S., Occhiodoro T., Bielicki J., Clements P.R., Hopwood J.J. Proc. Natl. Acad. Sci. U.S.A. 87:8531-8535(1990).
			[3] de Hostos E.L., Schilling J., Grossman A.R. Mol. Gen. Genet. 218:229-239(1989).
			[4] Selmer T., Hallmann A., Schmidt B., Sumper M., von Figura K. Eur. J. Biochem. 238:341-345(1996).
Sulfate_transp	PDOC00870	Sulfate transporters signature	A number of proteins involved in the transport of sulfate across a membrane as well as some yet uncharacterized proteins have been shown [1,2] to be evolutionary related. These proteins are:

		1	012
fam F	rosite	Full Name	Description
fam F	Prosite:	Full Name	- Neurospora crassa sulfate permease II (gene cys-14) Yeast sulfate permeases (genes SUL1 and SUL2) Rat sulfate anion transporter 1 (SAT-1) Mammalian DTDST, a probable sulfate transporter which, in Human, is involved in the genetic disease, diastrophic dysplasia (DTD) Sulfate transporters 1, 2 and 3 from the legume Stylosanthes hamata. - Human pendrin (gene PDS), which is involved in a number of hearing loss genetic diseases Human protein DRA (Down-Regulated in Adenoma) Soybean early nodulin 70 Escherichia coli hypothetical protein ychM.
			- Caenorhabditis elegans hypothetical protein F41D9.5. As expected by their transport function, these proteins are highly hydrophobic and seem to contain about 12 transmembrane domains. The best conserved region seems to be located in the second transmembrane region and is used as a signature pattern.
			Description of pattern(s) and/or profile(s) Consensus pattern [PAV]-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVFYA]-[LIVST]-[YI]- x(3)-[GA]-[GST]-S-[KR] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Sandal N.N., Marcker K.A. Trends Biochem. Sci. 19:19-19(1994). [2] Smith F.W., Hawkesford M.J., Prosser I.M., Clarkson D.T.
Synuclein		Synuclein	Mol. Gen. Genet. 247:709-715(1995). Accession number: PF01387 Definition: Synuclein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1] Gathering cutoffs: 25 25 Trusted cutoffs: 197.80 197.80 Noise cutoffs: -33.80 -33.80 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 98424410 Reference Author: Lavedan C; Reference Location: Genome Res 1998;8:871-880. INTERPRO; IPR001058; There are three types of synucleins in humans, these Comment: are called alpha, beta and gamma. Alpha synuclein has Comment: been found mutated in families with autosomal dominant Comment: Parkinson's disease. A peptide of alpha
			synuclein has Comment: also been found in amyloid plaques in Alzheimer's Comment: patients. Number of members: 12





Yam	Prosite	Full Name	Description
EA	PDOC00479	TEA domain signature	The TEA domain [1,E1] is a DNA-binding region of about 66 to
			68 amino acids
			which has been found in the N-terminal section of the
			following nuclear
			regulatory proteins:
			- Mammalian enhancer factor TEF-1. TEF-1 can bind to two distinct sequences
			in the SV40 enhancer and is a transcriptional activator.
			- Mammalian TEF-3, TEF-4 and TEF-5 [2], putative
			transcriptional activators
			highly similar to TEF-1.
			- Drosophila scalloped protein (gene sd), a probable
			transcription factor
			that functions in the regulation of cell-specific gene expression
			during
			Drosophila development, particularly in the differentiation of the
			nervous
			system [3].
			- Emericella nidulans regulatory protein abaA. AbaA is
			involved in the
			regulation of conidiation (asexual spore); its expression lead to the
			cessation of vegetative growth.
			- Yeast trans-acting factor TEC1. TEC1 is involved in the
			activation of the
			Ty1 retrotransposon.
			- Caenorhabditis elegans hypothetical protein F28B12.2.
			As a signature pattern, we have used positions 39 to 67 of the
			TEA domain.
			Description of pattern(s) and/or profile(s)
			Consensus pattern G-R-N-E-L-I-x(2)-Y-I-x(3)-[TC]-x(3)-R-T-
			[RK](2)-Q-[LIVM]- S-S-H-[LIVM]-Q-V
			Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update
			November 1997 / Pattern and text revised.
			References
			[1]
			Buerglin T.R.
			Cell 66:11-12(1991).
			[2]
			Jacquemin P., Hwang JJ., Martial J.A., Dolle P., Davidson I. J. Biol. Chem. 271:21775-21785(1996).
			l 31
			[3] Campbell S.D., Inamdar M., Rodrigues V., Raghavan V.,
			Palazzolo M., Chovnick A.
			Genes Dev. 6:367-379(1992).
			delica Bav. 6.667 676(1562).
			[E1]
			http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C002
TGT		Queuine tRNA-	Accession number: PF01702 .
		ribosyltransferase	Definition: Queuine tRNA-ribosyltransferase
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_1643 (release 4.1)
			Gathering cutoffs: -132 -132
			Trusted cutoffs: -110.00 -110.00
			Noise cutoffs: -155.40 -155.40
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 96256303 Reference Title: Crystal structure of tRNA-guanine
	1		

			U 1 4	
Pfam	Prosite	Full Name	Description	100 · 1 · 1 · 1 · 1 · 1
			Reference Title:	modification by base exchange.
			Reference Author: Reference Location:	Romier C, Reuter K, Suck D, Ficner R; EMBO J 1996;15:2850-2857.
			Reference Number:	[2]
			Reference Medline:	93287116
			Reference Title:	tRNA-guanine transglycosylase from
			Escherichia coli.	9
			Reference Title:	Overexpression, purification and quaternary
			structure.	0 1 04 1/ 1 1/4 01 0:
			Reference Author:	Garcia GA, Koch KA, Chong S; J Mol Biol 1993;231:489-497.
			Reference Location: Database Reference:	
	İ		PDBSUM]	SCOT, Thus, ia, [COOL TOOA][CATT
			Database Reference	INTERPRO; IPR002616;
			Database Reference	PDB; 1efz A; 138; 379;
			Database Reference	PDB; 1enu A; 138; 379;
			Database Reference	PDB; 1pud ; 138; 379;
			Database Reference	PDB; 1wkd ; 138; 379;
			Database Reference	PDB; 1wke ; 138; 379;
			Database Reference	PDB; 1wkf; 138; 379;
			Database reference:	PFAMB; PB037884; This is a family of queuine tRNA-
			Comment: ribosyltransferases	Tho is a family of quodino ti five-
			Comment:	EC:2.4.2.29, also known as tRNA-guanine
			transglycosylase	
			Comment:	and guanine insertion enzyme.
			Comment:	Queuine tRNA-ribosyltransferase modifies
			tRNAs for asparagine	
			Comment:	aspartic acid, histidine and tyrosine with
			queuine.	It catalyses the exchange of guanine-34 at
	ľ		Comment: the wobble position v	
			Comment:	7-aminomethyl-7-deazaguanine, and the
			addition of a cyclope	
			Comment:	moiety to 7-aminomethyl-7-deazaguanine-
			34 tRNA; giving a hy	permodified
			Comment:	base queuine in the wobble position [1,2].
			Comment:	The aligned region contains a zinc binding
			motif C-x-C-x2-C-x29 Comment:	and important tRNA and 7-aminomethyl-
			7deazaguanine bindi	
			Number of members	
				DE01010
Thi4		Thi4 family	Accession number: Definition:	PF01946
				hi4 family nright A, Ouzounis C, Bateman A
			Alignment method of	
			Source of seed mem	
		1	Gathering cutoffs:	25 25
			Trusted cutoffs:	526.80 526.80
			Noise cutoffs:	-105.00 -105.00
			HMM build command	d line: hmmbuild -F HMM SEED d line: hmmcalibrateseed 0 HMM
			Reference Number:	Inne: ninmcalibrateseed o Filvilvi [1]
			Reference Medline:	95050223
			Reference Title:	Cloning, nucleotide sequence, and
			regulation of	, , , , , , , , , , , , , , , , , , ,
			Reference Title:	Schizosaccharomyces pombe thi4, a
			thiamine biosynthetic	
			Reference Title:	gene.
			Reference Author: Reference Location:	Zurlinden A, Schweingruber ME; J Bacteriol 1994;176:6631-6635.
			Database Reference	
			Comment:	This family includes Swiss:P32318 a
			putative thiamine bid	
			Comment:	enzyme.
			Number of members	s: 14
ThiC		ThiC family	Accession number:	PF01964
1			Definition:	ThiC family
			Author: E	Enright A, Ouzounis C, Bateman A
			Alignment method o	
			Source of seed men Gathering cutoffs:	nbers: Enright A 25 25



		1	015
Yam	Prosite	Full Name	Description.
10011			Trusted cutoffs: 1047.20 1047.20
	-		Noise cutoffs: -338.20 -338.20
		İ	HMM build command line: hmmbuild -F HMM SEED hmmcalibrateseed 0 HMM
		I	
			Reference Number: [1]
			Reference Medline: 93163063 Reference Title: Structural genes for thiamine biosynthetic
			enzymes Reference Title: (thiCEFGH) in Escherichia coli K-12.
		1	
			V, Begley TP;
			Reference Location: J Bacteriol 1993;175:982-992.
			Reference Number: [2]
			Reference Medline: 99311269 Reference Title: Thiamin biosynthesis in prokaryotes.
			Reference Author: Begley 1P, Downs DM, Ealick SE,
			McLafferty FW, Van Loon AP, Reference Author: Taylor S, Campobasso N, Chiu HJ,
			Kinsland C, Reddick JJ, Xi
			Reference Author: J; Reference Location: Arch Microbiol 1999;171:293-300.
		1	
			Reference Number: [3]
	1		Reference Medline: 97284509 Reference Title: 97284509 Characterization of the Bacillus subtilis thir
	1		1,12,12,13
			operon Reference Title: involved in thiamine biosynthesis.
			The state of the s
			11.0101.0101.0001
			The state of the s
			Comment: ThiC is found within the tribanile
			biosynthesis operon. ThiC is Comment: involved in pyrimidine biosynthesis [2].
			- the substitution of the
	1		
		1	pyrophosphate of Comment: 2-methyl-4-amino-5-
			Comment: 2-methyl-4-amino-5-
			hydroxymethylpyrimidine pyrophosphate by Comment: 4-methyl-5-(beta-hydroxyethyl)thiazole
			phosphate to yield thiamine
			Comment: phosphate [3].
			Number of members: 12
			Accession number: PF01965
ThiJ		ThiJ/PfpI family	Acceptati training at
		l	- · · · · · · · · · · · · · · · · · · ·
			Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw
			Source of seed members: Enright A
		\	
			Gathering cutoffs: -40.2 -40.2 Trusted cutoffs: -40.20 -40.20
	1	1	
			HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
	1		Reference Medline: 97039868 Reference Title: The thiJ locus and its relation to
		1	
			phosphorylation of Reference Title: hydroxymethylpyrimidine in Escherichia
		1	coli.
			Reference Author: Mizote T, Tsuda M, Nakazawa T,
			Nakayama H;
			Nakayama H; Reference Location: Microbiology 1996;142:2969-2974.
			Nakayama H; Reference Location: Microbiology 1996;142:2969-2974. Reference Number: [2]
			Nakayama H; Reference Location: Microbiology 1996;142:2969-2974. Reference Number: [2] Reference Medline: 96196168
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: Nakayama H; Microbiology 1996;142:2969-2974. Microbiology 1996;142:2969-2974. Microbiology 1996;142:2969-2974. Sequence, expression in Escherichia co
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: And analysis of Microbiology 1996;142:2969-2974. Microbiology 1996;142:2969-2974. Microbiology 1996;142:2969-2974. Sequence, expression in Escherichia co
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: And analysis of Reference Title: Reference Title: And analysis of Reference Title: And analysis of Reference Title: And analysis of Reference Title: And analysis of Reference Title: And analysis of Reference Title: And analysis of Reference Title: And analysis of Reference Title: And analysis of Reference Location: Anicrobiology 1996;142:2969-2974. Bellion 1996;142:2969-2974. Bellion 1996;142:2969-2974. Reference Number: [2] Set 1996;142:2969-2974. Reference Number: [2] Reference Medline: Reference Medline: Reference Title: Anicrobiology 1996;142:2969-2974. Reference Number: Reference Medline: Reference Title: Anicrobiology 1996;142:2969-2974. Reference Medline: Reference Medline: Reference Title: Anicrobiology 1996;142:2969-2974. Reference Medline: Reference Title: Anicrobiology 1996;142:2969-2974. Reference Medline: Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Medline: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiology 1996;142:2969-2974. Reference Title: Anicrobiol
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: and analysis of Reference Title: protease (Pfpl) Microbiology 1996;142:2969-2974. [2] 96196168 Sequence, expression in Escherichia col the gene encoding a novel intracellular
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: and analysis of Reference Title: protease (PfpI) Reference Title: from the hyperthermophilic archaeon
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: and analysis of Reference Title: protease (Pfpl) Reference Title: Processes in the gene encoding a novel intracellular from the hyperthermophilic archaeon
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: and analysis of Reference Title: protease (Pfp!) Reference Title: Pyrococcus furiosus. Reference Author: Microbiology 1996;142:2969-2974. [2] 96196168 Sequence, expression in Escherichia col the gene encoding a novel intracellular from the hyperthermophilic archaeon Halio SB, Blumentals II, Short SA, Mer
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: and analysis of Reference Title: protease (Pfpl) Reference Title: Pyrococcus furiossus. Reference Author: Microbiology 1996;142:2969-2974. [2] 96196168 Sequence, expression in Escherichia col the gene encoding a novel intracellular from the hyperthermophilic archaeon Halio SB, Blumentals II, Short SA, Mer
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: and analysis of Reference Title: protease (Pfpl) Reference Title: Pyrococcus furiosus. Reference Author: BM, Kelly RM; Reference Location: Nicrobiology 1996;142:2969-2974. [2] 96196168 Sequence, expression in Escherichia col the gene encoding a novel intracellular from the hyperthermophilic archaeon Halio SB, Blumentals II, Short SA, Mer
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: and analysis of Reference Title: protease (Pfpl) Reference Title: Pyrococcus furiosus. Reference Author: BM, Kelly RM; Reference Location: Database Reference Microbiology 1996;142:2969-2974. [2] 96196168 Sequence, expression in Escherichia col the gene encoding a novel intracellular from the hyperthermophilic archaeon Halio SB, Blumentals II, Short SA, Mer
			Nakayama H; Reference Location: Reference Number: Reference Medline: Reference Title: and analysis of Reference Title: protease (Pfpl) Reference Title: Pyrococcus furiosus. Reference Author: BM, Kelly RM; Reference Location: Nicrobiology 1996;142:2969-2974. [2] 96196168 Sequence, expression in Escherichia collaboration in the gene encoding a novel intracellular from the hyperthermophilic archaeon Halio SB, Blumentals II, Short SA, Merosali, SB, Blumentals II, SB, SB, Blumentals II, SB, SB, SB, SB, SB, SB, SB, SB, SB, SB

		1	016
Pfam	Prosite	Full Name	Description Database reference: PFAMB; PB041784; Comment: This family includes ThiJ a thiamine biosynthesis Comment: enzyme [1] that catalyses the phosphorylation of Comment: hydroxymethylpyrimidine (HMP) to HMP monophosphate EC:2.7.1.49. Comment: The family also includes a the protease Pfpl Swiss:Q51732 [2]. Number of members: 34
Thr_dehydrat_C		C-terminal domain of Threonine dehydratase	Accession number: PF00585 Definition: C-terminal domain of Threonine dehydratase Previous Pfam IDs: Thr_dehydratase_C; Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Bateman A Gathering cutoffs: 25 25 Trusted cutoffs: 99.90 51.30 Noise cutoffs: -1.10 -1.10 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 98230745 Reference Title: Structure and control of pyridoxal phosphate dependent Reference Title: allosteric threonine deaminase. Gallagher DT, Gilliland GL, Xiao G, Zondlo J, Fisher KE, Reference Author: Chinchilla D, Eisenstein E; Reference Author: Structure 1998;6:465-475. Database Reference: SCOP; 1tdj; fa; [SCOP-USA][CATH-PDBSUM] Database Reference Database Reference Database Reference Database Reference Database Reference Comment: -I- Threonine dehydratases PALP all contain a carboxy Comment: terminal region. This region may have a regulatory role. Comment: Some members contain two copies of this region. Number of members: 30
thymidylat_synt	PDOC00086	Thymidylate synthase active site	Thymidylate synthase (EC 2.1.1.45) [1,2] catalyzes the reductive methylation of dUMP to dTMP with concomitant conversion of 5,10-methylenetetrahydrofolate to dihydrofolate. Thymidylate synthase plays an essential role in DNA synthesis and is an important target for certain chemotherapeutic drugs. Thymidylate synthase is an enzyme of about 30 to 35 Kd in most species except in protozoan and plants where it exists as a bifunctional enzyme that includes a dihydrofolate reductase domain. A cysteine residue is involved in the catalytic mechanism (it covalently binds the 5,6-dihydro-dUMP intermediate). The sequence around the active site of this enzyme is conserved from phages to vertebrates. Description of pattern(s) and/or profile(s) Consensus pattern R-x(2)-[LIVM]-x(3)-[FW]-[QN]-x(8,9)-[LV]-x-P-C-[HAVM]- x(3)-[QMT]-[FYW]-x-[LV] [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.

Pfam	Prosite	Full Name	O1 / Description
Pfam	riosite	Full Name	Last update
			November 1997 / Pattern and text revised.
			References
			[1]
			Benkovic S.J.
			Annu. Rev. Biochem. 49:227-251(1980).
			[2]
			Ross P., O'Gara F., Condon S.
			Appl. Environ. Microbiol. 56:2156-2163(1990).
Top6A		Type II DNA	Accession number: PF01962
		topoisomerase	Definition: Type II DNA topoisomerase
			Author: Enright A, Ouzounis C, Bateman A
			Alignment method of seed: Clustalw Source of seed members: Enright A
			Gathering cutoffs: -99 -99
			Trusted cutoffs: -40.40 -40.40
			Noise cutoffs: -158.40 -158.40
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			Reference Medline: 97238688
			Reference Title: An atypical topoisomerase II from Archaea
			with implications
			Reference Title: for meiotic recombination [see comments]
			Reference Author: Bergerat A, de Massy B, Gadelle D,
			Varoutas PC, Nicolas A, Reference Author: Forterre P;
			Reference Location: Nature 1997;386:414-417.
			Database Reference: SCOP; 1d3y; fa; [SCOP-USA][CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR002815;
			Database Reference PDB; 1d3y A; 77; 363; PDB; 1d3y B; 77; 363; PDB; 1d3y B; 77; 363;
			Database Reference PDB; 1d3y B; 77; 363; Comment: Members of this family are the A subunit
			from type II DNA
			Comment: topoisomerases. Type II DNA
			topoisomerases catalyse the relaxation
			Comment: of DNA supercoiling by causing transient
			double strand breaks.
			Comment: The family includes topoisomerase VI subunit A from archaebacteria
			Comment: Swiss:Q57815 EC:5.99.1.3 and SPO11
			from yeast Swiss:P23179.
			Comment: A conserved tyrosine is thought to be
			involved in breaking the
	1		Comment: double stranded DNA [1]. Number of members: 9
			Trumber of members.
Topoisom bac	PDOC00333	Prokaryotic DNA	DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the
		topoisomerase I active	two types of
		site	enzyme that catalyze the interconversion of topological DNA
			isomers. Type I
			topoisomerases act by catalyzing the transient breakage of DNA, one strand at
			a time, and the subsequent rejoining of the strands. When a
			prokaryotic type I
	}		topoisomerase breaks a DNA backbone bond, it simultaneously
:			forms a protein-
			DNA link where the hydroxyl group of a tyrosine residue is
			joined to a 5'- phosphate on DNA, at one end of the enzyme-severed DNA
			strand.
			Prokaryotic organisms, such as Escherichia coli, have two type I
			topoisomerase I (gene topA) and topoisomerase III
			isozymes: topoisomerase I (gene topA) and topoisomerase III (gene topB).
1			Eukaroytes also contain homologs of prokaryotic topoisomerase
			III.
			There are a number of conserved residues in the region around the active site

		1	018
Pfam	Prosite	Full Name	Description
			tyrosine; we used this region as a signature pattern.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [EQ]-x-L-Y-[DEQST]-x(3,12)-[LIV]-[ST]-Y-x-R-[ST]-[DEQS] [The second Y is the active site tyrosine] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References
			[1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).
			[2] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).
			[3] Bjornsti MA. Curr. Opin. Struct. Biol. 1:99-103(1991).
			[4] Roca J. Trends Biochem. Sci. 20:156-160(1995).
			[E1] http://ellington.pharm.arizona.edu/~bear/top/topo.html
toxin_3		long chain scorpion toxins	Accession number: PF00537 Definition: long chain scorpion toxins Author: Bateman A Alignment method of seed: Manual Source of seed members: Arne Elofsson. Gathering cutoffs: 25 25 Trusted cutoffs: 59.50 59.50 Noise cutoffs: -3.80 -3.80 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference: SCOP; 2sn3; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002061; Comment: -!- Scorpion toxins bind to sodium channels and inhibit the activation Comment: mechanisms of the channels, thereby blocking neuronal transmission. Number of members: 77
Translin		Translin family	Accession number: PF01997 Definition: Translin family Previous Pfam IDs: DUF130; Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 299.50 299.50 Noise cutoffs: -72.40 -72.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] 97165975 Isolation and characterization of a cDNA encoding a Reference Title: Translin-like protein, TRAX. Aoki K, Ishida R, Kasai M; Reference Author: Aoki K, Ishida R, Kasai M; Reference Location: Database Reference Comment: Members of this family include Translin Swiss:Q15631 that interacts Comment: with DNA and forms a ring around the DNA.





		1	019
Pfam	Prosite	Full Name	Description
			This family also includes Comment: Swiss:Q99598, that was found to interact with translin with yeast Comment: two-hybrid screen [1]. Number of members: 10
Transposase_19		Transposase 19	Members of this family are capable of in vitro and/or in vivo insertion of a donor polynucleotide into a target polynucleotide. Such biological activity is useful for inserting DNA into host genome, for example, for cloning purposes to generate a desired vector in vitro.
TRANSPOSASE_IS 30	PDOC00801	Transposases, IS30 family, signature	Autonomous mobile genetic elements such as transposon or insertion sequences (IS) encode an enzyme, called transposase, required for excising and inserting the mobile element. On the basis of sequence similarities, transposases can be grouped into various families. One of these families has been shown [1,2] to consist of transposases from the following elements: - Is30 from Escherichia coli Is1086 from Alcaligenes eutrophus Is1161 from Streptococcus salivarius Is4351 (Tn4551) from Bacteroides fragilis. These transposases are proteins of 340 to 380 amino acids. The best conserved region is located in their C-terminal section and is used as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern R-G-x(2)-E-N-x-N-G-[LIVM](2)-R-[QE]-[LIVMFY](2)-P-K Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / First entry. References [1] Dong Q., Sadouk A., van der Lelie D., Taghavi S., Ferhat A., Nuyten J.M., Borremans B., Mergeay M., Toussaint A. J. Bacteriol. 174:8133-8138(1992). [2] Giffard P.M., Rathsam C., Kwan E., Kwan D.W.L., Bunny K.L.,
			Koo SP., Jacques N.A. J. Gen. Microbiol. 139:913-920(1993).
Transthyretin	PDOC00617	Transthyretin signatures	Transthyretin (prealbumin) [1] is a thyroid hormone-binding protein that seems to transport thyroxine (T4) from the bloodstream to the brain. It is a protein of about 130 amino acids that assembles as a homotetramer and forms an internal channel that binds thyroxine. Transthyretin is mainly synthesized in the brain choroid plexus. In humans, variants of the protein are associated with distinct forms of amyloidosis. The sequence of transthyretin is highly conserved in vertebrates. A number of uncharacterized proteins also belong to this family: - Escherichia coli hypothetical protein yedX Bacillus subtilis hypothetical protein yunM.



Dfam	Prosite	Eul Nome	Description
Pfam	LIOSIIA	Full Name	Description Cooperhabilitie alegans hypothetical protein B00H10.3
			- Caenorhabditis elegans hypothetical protein R09H10.3. - Caenorhabditis elegans hypothetical protein ZK697.8.
			We selected two regions as signature patterns. The first located in the N-
			terminal extremity starts with a lysine known to be involved in binding T4.
			The second pattern is located in the C-terminal extremity.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [KH]-[IV]-L-[DN]-x(3)-G-x-P-A-x(2)-[IV]-x-[IV] [The K binds thyroxine] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern Y-[TH]-[IV]-[AP]-x(2)-L-S-[PQ]-[FYW]-[GS]-
			[FY]-[QS] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update July 1999 / Patterns and text revised.
			References [1] Schreiber G. Richardson S. I.
			Schreiber G., Richardson S.J. Comp. Biochem. Physiol. 116B:137-160(1997).
TRM		N2,N2- dimethylguanosine tRNA	Accession number: PF02005 Definition: N2,N2-dimethylguanosine tRNA
		methyltransferase	methyltransferase Author: Enright A, Ouzounis C, Bateman A
			Alignment method of seed: Clustalw Source of seed members: Enright A
			Gathering cutoffs: 25 25 Trusted cutoffs: 664.60 664.60
•			Noise cutoffs: -259.50 -259.50 HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1]
			Reference Medline: 98352211
			Reference Title: The tRNA(guanine-26,N2-N2) methyltransferase (Trm1) from
			Reference Title: the hyperthermophilic archaeon Pyrococcus furiosus:
			Reference Title: cloning, sequencing of the gene and its expression in
			Reference Title: Escherichia coli. Reference Author: Constantinesco F, Benachenhou N,
			Motorin Y, Grosjean H; Reference Location: Nucleic Acids Res 1998;26:3753-3761.
			Reference Number: [2] Reference Medline: 87260951
			Reference Title: Amino-terminal extension generated from an upstream AUG
			Reference Title: codon is not required for mitochondrial import of yeast
			Reference Title: N2,N2-dimethylguanosine- specific tRNA methyltransferase.
:			Reference Author: Ellis SR, Hopper AK, Martin NC; Reference Location: Proc Natl Acad Sci U S A 1987;84:5172- 5176.
			Database Reference INTERPRO; IPR002905; Database reference: PFAMB; PB041661;
			Comment: This enzyme EC:2.1.1.32 used S-AdoMet to methylate tRNA.
			Comment: The TRM1 gene of Saccharomyces cerevisiae is necessary for
			Comment: the N2,N2-dimethylguanosine modification of both mitochondrial
		L	Comment: and cytoplasmic tRNAs [1]. The enzyme is

Pfam	Prosite	M 10444-14	021
Plam	Prosite	Full Name	Description found in both
			Comment: eukaryotes and archaebacteria [2]
			Number of members: 10
			Traditibel of members.
tRNA bind		Putative tRNA binding	Accession number: PF01588
		Idomain	Definition: Putative tRNA binding domain
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B 482 (release 4.1)
			Gathering cutoffs: 20 20
			Trusted cutoffs: 22.30 22.30
			Noise cutoffs: 18.20 18.20
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 97306356
			Reference Title: Human tyrosyl-tRNA synthetase shares
			amino acid sequence Reference Title: homology with a putative cytokine.
			Reference Title: homology with a putative cytokine. Reference Author: Kleeman TA, Wei D, Simpson KL, First
			EA:
			Reference Location: J Biol Chem 1997;272:14420-14425.
			Reference Number: [2]
			Reference Medline: 97050848
			Reference Title: The yeast protein Arc1p binds to tRNA and
			functions as a
			Reference Title: cofactor for the methionyl-and glutamyl-
			tRNA synthetases.
			Reference Author: Simos G, Segref A, Fasiolo F, Hellmuth K,
			Shevchenko A,
			Reference Author: Mann M, Hurt EC;
			Reference Location: EMBO J 1996;15:5437-5448.
			Database Reference: SCOP; 1pys; fa; [SCOP-USA][CATH-
			PDBSUM]
			Database Reference INTERPRO; IPR002547;
			Database Reference PDB; 1570 B; 153; 247; Database Reference PDB; 1570 B; 153; 247; PDB; 1570 B; 153; 247;
			Database Reference PDB; 1b7y B; 153; 247; Database Reference PDB; 1eiy B; 153; 247;
			Database Reference PDB; 1pys B; 153; 247;
			Database reference: PFAMB; PB010015;
			Comment: This domain is found in prokaryotic
			methionyl-tRNA synthetases,
			Comment: prokaryotic phenylalanyl tRNA synthetases
			the yeast GU4 nucleic-binding
			Comment: protein (G4p1 or p42, ARC1) [2], human
			tyrosyl-tRNA synthetase [1],
			Comment: and endothelial-monocyte activating
(4)			polypeptide II.
			Comment: G4p1 binds specifically to tRNA form a
			complex with methionyl-tRNA
			Comment: synthetases [2]. In human tyrosyl-tRNA
			synthetase this domain may direct
-			Comment: tRNA to the active site of the enzyme [2].
			This domain may perform a Comment: common function in tRNA aminoacylation
Ì			[1].
			Number of members: 46
			Trainbor of members:
RNA-synt 2d	PDOC00363	Aminoacyl-transfer RNA	Aminoacyl-tRNA synthetases (EC 6.1.1) [1] are a group of
		synthetases class-II	enzymes which
		signatures	activate amino acids and transfer them to specific tRNA
			molecules as the first
			step in protein biosynthesis. In prokaryotic organisms there are
			at least
			twenty different types of aminoacyl-tRNA synthetases, one for
			each different
			amino acid. In eukaryotes there are generally two aminoacyl-
			tRNA synthetases
			for each different amino acid: one cytosolic form and a
			for each different amino acid: one cytosolic form and a mitochondrial form.
			for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are
			for each different amino acid: one cytosolic form and a mitochondrial form.



		1()22
Pfam P	rosite		Description The synthetases specific for alanine, asparagine, aspartic acid, glycine,
			histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common
			folding pattern in their catalytic domain for the binding of ATP and amino acid
			which is different to the Rossmann fold observed for the class I synthetases [7].
		4	Class-II tRNA synthetases do not share a high degree of similarity, however at
			least three conserved regions are present [2,5,8]. We have derived signature patterns from two of these regions.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE] Sequences known to belong to this class detected by the pattern the majority of class-II tRNA synthetases with the exception of those specific for alanine, glycine as well as bacterial histidine. Other sequence(s) detected in SWISS-PROT 43.
			Consensus pattern [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x- [LIVMSTAG]-[LIVMFY] Sequences known to belong to this class detected by the pattern the majority of class-II tRNA synthetases with the exception of those specific for serine and proline. Other sequence(s) detected in SWISS-PROT 161. Expert(s) to contact by email Cusack S. cusack@embl-grenoble.fr
			Last update July 1998 / Text revised. References [1]
			Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).
			[2] Delarue M., Moras D. BioEssays 15:675-687(1993).
			[3] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).
			[4] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).
			[5] Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).
			[6] Cusack S. Biochimie 75:1077-1081(1993).
			[7] Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347:249-255(1990).
			[8] Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).
trypsin F	PDOC00124	Serine proteases, trypsin family, active sites	The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-

		Τ,	023
Pfam	Prosite	Full Name	Description
			bonded to a histidine, which itself is hydrogen-bonded to a
			serine. The
			sequences in the vicinity of the active site serine and histidine residues are
			well conserved in this family of proteases [1]. A partial list of
			proteases
			known to belong to the trypsin family is shown below.
			- Acrosin.
			- Blood coagulation factors VII, IX, X, XI and XII, thrombin,
			plasminogen,
			and protein C.
			- Cathepsin G.
			- Chymotrypsins Complement components C1r, C1s, C2, and complement
			factors B, D and I.
			- Complement-activating component of RA-reactive factor.
			- Cytotoxic cell proteases (granzymes A to H).
			- Duodenase I.
			- Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin).
			- Enterokinase (EC 3.4.21.9) (enteropeptidase).
			- Hepatocyte growth factor activator Hepsin.
			- Glandular (tissue) kallikreins (including EGF-binding protein
	1		types A, B,
			and C, NGF-gamma chain, gamma-renin, prostate specific
			antigen (PSA) and
			tonin).
			- Plasma kallikrein.
			- Mast cell proteases (MCP) 1 (chymase) to 8 Myeloblastin (proteinase 3) (Wegener's autoantigen).
			- Plasminogen activators (urokinase-type, and tissue-type).
			- Trypsins I, II, III, and IV.
			- Tryptases.
			- Snake venom proteases such as ancrod, batroxobin,
		ļ	cerastobin, flavoxobin,
			and protein C activator Collagenase from common cattle grub and collagenolytic
			protease from
			Atlantic sand fiddler crab.
			- Apolipoprotein(a).
			- Blood fluke cercarial protease.
			- Drosophila trypsin like proteases: alpha, easter, snake-locus.
			Drosophila protease stubble (gene sb). Major mite fecal allergen Der p III.
			All the above proteins belong to family S1 in the classification of
Ì			peptidases
	-		[2,E1] and originate from eukaryotic species. It should be
			noted that
			bacterial proteases that belong to family S2A are similar
			lenough in the regions of the active site residues that they can be picked up by
			the same
			patterns. These proteases are listed below.
			Achromohacter lytique proteases I
			- Achromobacter lyticus protease I. - Lysobacter alpha-lytic protease.
	1		- Streptogrisin A and B (Streptomyces proteases A and B).
			- Streptomyces griseus glutamyl endopeptidase II.
			- Streptomyces fradiae proteases 1 and 2.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIVM]-[ST]-A-[STAG]-H-C [H is the active site
			residue] Sequences known to belong to this class detected by the pattern
			ALL, except for complement components C1r and C1s, pig
			plasminogen, bovine protein C, rodent urokinase, ancrod, gyroxin
1			and two insect trypsins.
1			Other sequence(s) detected in SWISS-PROT 14.



			024
Pfam	Prosite	Full Name	Description Consensus pattern [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]- [LIVMFYWH]-[LIVMFYSTANQH] [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for 18 different proteases which have lost the first conserved glycine. Other sequence(s) detected in SWISS-PROT H.influenzae protease HAP which belongs to family S6 and 3 other proteins. Note if a protein includes both the serine and the histidine active site signatures, the probability of it being a trypsin family serine protease is 100% Last update November 1997 / Text revised. References [1] Brenner S. Nature 334:528-530(1988). [2] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994). [E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt
TYA		TYA transposon protein	Accession number: PF01021 Definition: TYA transposon protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_90 (release 3.0) Gathering cutoffs: 15 15 Trusted cutoffs: 18.00 18.00 Noise cutoffs: 13.70 13.70 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 97404699 Reference Title: retrotransposon virus-like particles. Reference Author: NR, Butcher SJ, Reference Author: NR, Butcher SJ, Reference Location: Database Reference Comment: Ty are yeast transposons. A 5.7kb transcript codes Comment: for p3 a fusion protein of TYA and TYB. The TYA Comment: protein is analogous to the gag protein of retroviruses. Comment: TYA a is cleaved to form 46kd protein which can form Comment: mature virion like particles [1].
tyrosinase	PDOC00398	Tyrosinase signatures	Tyrosinase (EC 1.14.18.1) [1] is a copper monooxygenases that catalyzes the hydroxylation of monophenols and the oxidation of o-diphenols to o-quinols. This enzyme, found in prokaryotes as well as in eukaryotes, is involved in the formation of pigments such as melanins and other polyphenolic compounds. Tyrosinase binds two copper ions (CuA and CuB). Each of the two copper ion has been shown [2] to be bound by three conserved histidines residues. The regions around these copper-binding ligands are well conserved and also shared by some hemocyanins, which are copper-containing oxygen carriers from the hemolymph of



		1	025
Pfam	Prosite	Full Name	Description
			many molluscs and arthropods [3,4].
			At least two proteins related to tyrosinase are known to exist in mammals:
			- TRP-1 (TYRP1) [5], which is responsible for the conversion of 5,6-dihydro-
			xyindole-2-carboxylic acid (DHICA) to indole-5,6-quinone-2-carboxylic acid.
			- TRP-2 (TYRP2) [6], which is the melanogenic enzyme DOPAchrome tautomerase
			(EC 5.3.3.12) that catalyzes the conversion of DOPAchrome to DHICA. TRP-2
			differs from tyrosinases and TRP-1 in that it binds two zinc ions instead of copper [7].
			Other proteins that belong to this family are:
			- Plants polyphenol oxidases (PPO) (EC 1.10.3.1) which catalyze the oxidation
			of mono- and o-diphenols to o-diquinones [8] Caenorhabditis elegans hypothetical protein C02C2.1.
			We have derived two signature patterns for tyrosinase and related proteins.
			The first one contains two of the histidines that bind CuA, and is located in
			the N-terminal section of tyrosinase. The second pattern contains a histidine
			that binds CuB, that pattern is located in the central section of the enzyme.
			Description of pattern(s) and/or profile(s)
			Consensus pattern H-x(4,5)-F-[LIVMFTP]-x-[FW]-H-R-x(2)-[LVM]-x(3)-E [The two H's are copper ligands] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern D-P-x-F-[LIVMFYW]-x(2)-H-x(3)-D [H is a copper ligand]
			Sequences known to belong to this class detected by the pattern ALL the tyrosinases as well as all the hemocyanins. Other sequence(s) detected in SWISS-PROT NONE.
			Last update December 1999 / Patterns and text revised. References
			[1] Lerch K. Prog. Clin. Biol. Res. 256:85-98(1988).
			[2]
			Jackman M.P., Hajnal A., Lerch K. Biochem. J. 274:707-713(1991).
			[3] Linzen B. Naturwissenschaften 76:206-211(1989).
			[4] Lang W.H., van Holde K.E. Proc. Natl. Acad. Sci. U.S.A. 88:244-248(1991).
			[5] Kobayashi T., Urabe K., Winder A., Jimenez-Cervantes C., Imokawa G., Brewington T., Solano F., Garcia-Borron J.C., Hearing V.J. EMBO J. 13:5818-5825(1994).
			[6]



D.C.	Im. a	F . 1 . 1	in
Pfam	Prosite	Full Name	Description Jackson I.J., Chambers D.M., Tsukamoto K., Copeland N.G., Gilbert D.J., Jenkins N.A., Hearing V. EMBO J. 11:527-535(1992). [7] Solano F., Martinez-Liarte J.H., Jimenez-Cervantes C., Garcia-Borron J.C., Lozano J.A. Biochem. Biophys. Res. Commun. 204:1243-1250(1994). [8] Cary J.W., Lax A.R., Flurkey W.H. Plant Mol. Biol. 20:245-253(1992).
UbiA	PDOC00727	UbiA prenyltransferase family signature	The following prenyltransferases are evolutionary related [1,2]: - Bacterial 4-hydroxybenzoate octaprenyltransferase (gene ubiA) Yeast mitochondrial para-hydroxybenzoate polyprenyltransferase (gene COQ2) Protoheme IX farnesyltransferase (heme O synthase) from yeast and mammals (gene COX10) and from bacteria (genes cyoE or ctaB). These proteins probably contain seven transmembrane segments. The best conserved region is located in a loop between the second and third of these segments and we used it as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern N-x(3)-[DEH]-x(2)-[LIMF]-D-x(2)-[VM]-x-R- [ST]-x(2)-R-x(4)- G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Melzer M., Heide L. Biochim. Biophys. Acta 1212:93-102(1994). [2] Mogi T., Saiki K., Anraku Y. Mol. Microbiol. 14:391-398(1994).
Ubie_methyltran	PDOC00911	ubiE/COQ5 methyltransferase family signatures	The following methyltransferases have been shown [1] to share regions of similarities: - Escherichia coli ubiE, which is involved in both ubiquinone and menaquinone biosynthesis and which catalyzes the S-adenosylmethionine dependent methylation of 2-polyprenyl-6-methoxy-1,4-benzoquinol into 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinol and of demethylmenaquinol into menaquinol Yeast COQ5, a ubiquinone biosynthesis methlytransferase Bacillus subtilis spore germination protein C2 (gene: gercB or gerC2), a probable menaquinone biosynthesis methlytransferase Lactococcus lactis gerC2 homolog Caenorhabditis elegans hypothetical protein ZK652.9 Leishmania donovani amastigote-specific protein A41. These are hydrophilic proteins of about 30 Kd (except for ZK652.9 which is 65 Kd). They can be picked up in the database by the following patterns.

			027
Pfam	Prosite	Full Name	Description
	*		Description of pattern(s) and/or profile(s)
			Consensus pattern Y-D-x-M-N-x(2)-[LIVM]-S-x(3)-H-x(2)-W Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Consensus pattern R-V-[LIVM]-K-[PV]-[GM]-G-x-[LIVMF]-x(2)-
			[LIVM]-E-x-S Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update
			December 1999 / Pattern and text revised. References
			Lee P.T., Hsu A.Y., Ha H.T., Clarke C.F. J. Bacteriol. 179:1748-1754(1997).
ubiquitin	PDOC00271	Ubiquitin domain signature and profile	Ubiquitin [1,2,3] is a protein of seventy six amino acid residues, found in all eukaryotic cells and whose sequence is extremely well
			conserved from protozoan to vertebrates. It plays a key role in a variety of
			cellular processes, such as ATP-dependent selective degradation of cellular proteins,
			maintenance of chromatin structure, regulation of gene expression, stress response and ribosome biogenesis.
			In most species, there are many genes coding for ubiquitin.
			However they can be classified into two classes. The first class produces polyubiquitin
			molecules consisting of exact head to tail repeats of ubiquitin. The number of
			repeats is variable (up to twelve in a Xenopus gene). In the majority of polyubiquitin precursors, there is a final amino-acid after the last
· po			repeat. The second class of genes produces precursor proteins
			consisting of a single copy of ubiquitin fused to a C-terminal extension protein (CEP). There are two
			types of CEP proteins and both seem to be ribosomal proteins.
			Ubiquitin is a globular protein, the last four C-terminal residues (Leu-Arg-
			Gly-Gly) extending from the compact structure to form a 'tail', important for its function. The latter is mediated by the covalent conjugation of
			ubiquitin to target proteins, by an isopeptide linkage between the C-
			terminal glycine and the epsilon amino group of lysine residues in the target proteins.
			There are a number of proteins which are evolutionary related to ubiquitin:
			- Ubiquitin-like proteins from baculoviruses as well as in some strains of
			bovine viral diarrhea viruses (BVDV). These proteins are highly similar to
			their eukaryotic counterparts Mammalian protein GDX [4]. GDX is composed of two domains, a N-terminal
			ubiquitin-like domain of 74 residues and a C-terminal domain of 83 residues
			with some similarity with the thyroglobulin hormonogenic site Mammalian protein FAU [5]. FAU is a fusion protein which



	1	028
Pfam Prosite	Full Name	Description
		consist of a
		N-terminal ubiquitin-like protein of 74 residues fused to
		ribosomal protein S30.
		- Mouse protein NEDD-8 [6], a ubiquitin-like protein of 81
		residues.
		- Human protein BAT3, a large fusion protein of 1132 residues
		that contains a
		N-terminal ubiquitin-like domain.
		- Caenorhabditis elegans protein ubl-1 [7]. Ubl-1 is a fusion
		protein which
		consist of a N-terminal ubiquitin-like protein of 70 residues fused to
		ribosomal protein S27A.
		- Yeast DNA repair protein RAD23 [8]. RAD23 contains a N-
		terminal domain that
		seems to be distantly, yet significantly, related to ubiquitin.
		- Mammalian RAD23-related proteins RAD23A and RAD23B.
		- Mammalian BCL-2 binding athanogene-1 (BAG-1). BAG-1 is
		a protein of 274
		residues that contains a central ubiquitin-like domain Human spliceosome associated protein 114 (SAP 114 or
		SF3A120).
		- Yeast protein DSK2, a protein involved in spindle pole body
		duplication and
		which contains a N-terminal ubiquitin-like domain.
1		- Human protein CKAP1/TFCB, Schizosaccharomyces pombe
,		protein alp11 and
		Caenorhabditis elegans hypothetical protein F53F4.3. These
		proteins contain a N-terminal ubiquitin domain and a C-terminal CAP-Gly
		domain (see
		<pdoc00660>).</pdoc00660>
		- Schizosaccharomyces pombe hypothetical protein
		SpAC26A3.16. This protein
		contains a N-terminal ubiquitin domain.
		V OMTO
		- Yeast protein SMT3.
		- Human ubiquitin-like proteins SMT3A and SMT3B Human ubiquitin-like protein SMT3C (also known as PIC1; Ubl1,
		Sumo-1; Gmp-1
		or Sentrin). This protein is involved in targeting ranGAP1 to the
		nuclear
		pore complex protein ranBP2.
		- SMT3-like proteins in plants and Caenorhabditis elegans.
		To identify ubiquitin and related proteins we have developed a
		pattern based
		on conserved positions in the central section of the sequence. A
		profile was
		also developed that spans the complete length of the ubiquitin
		domain.
		Description of pattern(s) and/or profile(s)
		Consensus pattern K-x(2)-[LIVM]-x-[DESAK]-x(3)-[LIVM]-[PA]-
		x(3)-Q-x-[LIVM]- [LIVMC]-[LIVMFY]-x-G-x(4)-[DE]
		Sequences known to belong to this class detected by the pattern
		ALL, except for the RAD23 and SMT3 subfamilies, BAG-1 and
		SAP 114.
		Other sequence(s) detected in SWISS-PROT NONE.
		Sequences known to belong to this class detected by the profile
		ALL.
		Other sequence(s) detected in SWISS-PROT NONE.
		Note this documentation entry is linked to both a signature pattern
		and a profile. As the profile is much more sensitive than the
		pattern, you should use it if you have access to the necessary
	1	software tools to do so.
		Last update
1	1	July 1998 / Text revised.



		1	029
Pfam	Prosite	Full Name	Description
			Bio/Technology 8:209-215(1990). References
			[1]
			Jentsch S., Seufert W., Hauser HP. Biochim. Biophys. Acta 1089:127-139(1991).
			Biochini. Biophys. Asia 1005.127 105(1051).
			[2]
			Monia B.P., Ecker D.J., Croke S.T.
			1 2)
			[3] Finley D., Varshavsky A.
			Trends Biochem. Sci. 10:343-347(1985).
			· ·
			[4]
			Filippi M., Tribioli C., Toniolo D. Genomics 7:453-457(1990).
			denomics 7.400-407 (1000).
		•	[5]
			Olvera J., Wool I.G.
			J. Biol. Chem. 268:17967-17974(1993).
			(6)
			[6] Kumar S., Yoshida Y., Noda M.
			Biochem, Biophys. Res. Commun. 195:393-399(1993).
			[7]
			Jones D., Candido E.P. J. Biol. Chem. 268:19545-19551(1993).
			0. Biol. Chem. 208.19040-19001(1990).
			[8]
			Melnick L., Sherman F.
			J. Mol. Biol. 233:372-388(1993).
LIDE0004	DD0000004	Linebarasterized protein	The following uncharacterized proteins have been shown [1] to
UPF0004	PDOC00984	Uncharacterized protein family UPF0004	share regions of
		signature	similarities:
		3	
			- Escherichia coli hypothetical protein yliG.
			- Escherichia coli hypothetical protein yleA and HI0019, the
			corresponding Haemophilus influenzae protein.
			- Bacillus subtilis hypothetical protein yqeV.
			- Helicobacter pylori hypothetical protein HP0269.
			- Helicobacter pylori hypothetical protein HP0285.
			- Mycoplasma iowae hypothetical protein in 16S RNA 5'region. - Mycobacterium tuberculosis hypothetical protein Rv2733c.
			- Rickettsia prowazekii hypothetical protein RP416.
			- Rickettsia prowazekii hypothetical protein RP808.
			- Synechocystis strain PCC 6803 hypothetical protein slr0082.
		1	- Synechocystis strain PCC 6803 hypothetical protein sll0996.
			- Methanococcus jannaschii hypothetical protein MJ0865. - Methanococcus jannaschii hypothetical protein MJ0867.
			- Caenorhabditis elegans hypothetical protein F25B5.5.
			7,
			The size of these proteins range from 47 to 61 Kd. They contain
			six conserved
			cysteines, three of which are clustered in a region that can be used as a
			signature pattern.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [LIVM]-x-[LIVMT]-x(2)-G-C-x(3)-C-[STAN]-
			[FY]-C-x-[LIVMT]- x(4)-G
			Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT 2.
			Last update
			December 1999 / Pattern and text revised.
	1	1	References
1			r 41
			[1] Bairoch A.



	1		030
Pfam	Prosite	Full Name	Description Unpublished observations (1997).
UPF0013		Uncharacterized membrane protein family UPF0013	Accession number: PF01554 Definition: Uncharacterized membrane protein family UPF0013 Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_163 (release 4.0) Gathering cutoffs: -26 -26 Trusted cutoffs: -16.10 -16.10 Noise cutoffs: -36.70 -36.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference: URL; http://www.expasy.ch/cgi-bin/lists?upflist.bt; Database Reference: URL; http://www.expasy.ch/cgi-bin/lists?upflist.bt; Database Reference: INTERPRO; IPR002528; Database reference: PFAMB; PB041103; These proteins are integral membrane proteins of unknown Comment: function. Number of members: 47
UPF0019	PDOC00949	Uncharacterized protein family UPF0019 signature	The following uncharacterized proteins have been shown [1,2] to be highly similar: - Yeast protein SNZ1, which may be involved in growth arrest and cellular response to nutrient limitation. - Yeast chromosome VI hypothetical protein YFL059w. - Yeast chromosome XIV hypothetical protein YNL333w. - Fission yeast hypothetical protein SpAC29B12.04. - Hevea brasiliensis ethylene-inducible protein HEVER. - Stellaria longipes hypothetical protein H47. - Bacillus subtilis hypothetical protein yaaD. - Haemophilus influenzae hypothetical protein MICL581.12c. - Mycobacterium leprae hypothetical protein MICL581.12c. - Mycobacterium tuberculosis hypothetical protein MtCY1A10.27. - Archaeoglobus fulgidus hypothetical protein MJ0677. - Methanococcus yannaschii hypothetical protein MJ0677. - Methanococcus vannielii hypothetical protein in tRNA/5S rRNA gene cluster. - Methanobacterium thermoautotrophicum hypothetical protein Mth666. These are hydrophilic proteins of about 32 Kd. They can be picked up in the database by the following pattern. Description of pattern(s) and/or profile(s) Consensus pattern L-P-V-[VT]-[NQL]-F-[AT]-A-G-G-[LIV]-A-T-P-A-D-A-A-[LM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / Pattern and text revised. References [1] Sivasubramaniam S., Vanniasingham V.M., Tan C.T., Chua N.H. Plant Mol. Biol. 29:173-178(1995).
UPF0047	PDOC01018	Uncharacterized protein family UPF0047 signature	Braun E.L., Fuge E.K., Padilla P.A., Werner-Washburne M. J. Bacteriol. 178:6865-6872(1996). The following uncharacterized proteins have been shown [1] to be highly similar:
			- Bacillus subtilis hypothetical protein yugU.



			731
Mam	Prosite		Description - Escherichia coli hypothetical protein yjbQ. - Mycobacterium tuberculosis hypothetical protein MtCY9C4.12. - Synechocystis strain PCC 6803 hypothetical protein sll1880. - Archaeoglobus fulgidus hypothetical protein AF2050. - Methanococcus jannaschii hypothetical protein MJ1081. - Methanobacterium thermoautotrophicum hypothetical protein MTH771. - Fission yeast hypothetical protein SpAC4A8.02c. These are small proteins of 14 to 16 Kd. They can be picked up in the database by the following pattern. This pattern is located in the C-terminal part of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern S-X(2)-[LIV]-x-[LIV]-x(2)-G-x(4)-G-T-W-Q-x-[LIV] Sequences known to belong to this class detected by the pattern
			ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / First entry. References [1] Bairoch A. Unpublished observations (1998).
UPF0052		Uncharacterised protein family UPF0052	Accession number: PF01933 Definition: Uncharacterised protein family UPF0052 Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 263.90 263.90 Noise cutoffs: -134.40 -134.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference INTERPRO; IPR002882; Number of members: 12
UPF0057	PDOC01013	Uncharacterized protein family UPF0057 signature	The following uncharacterized proteins have been shown [1] to be evolutionary related: - Barley low-temperature induced protein blt101 Lophorium elongatum salt-sress induced protein ESI3 Yeast hypothetical proteins YDL123w, YDR276c, YDR525Bw and YJL151c Caenorhabditis elegans hypothetical proteins F47B7.1, T23F2.3, T23F2.4, T23F2.5 and ZK632.10 Escherichia coli hypothetical protein yqaE Synechocystis strain PCC 6803 hypothetical protein ssr1169. These are small proteins of from 52 to 140 amino-acid resiudes that contains two transmembrane domains. As a signature pattern we selected a region that corresponds to the end of the first transmembrane helix.
			Description of pattern(s) and/or profile(s) Consensus pattern [LIV]-x-[STA]-[LIVF](3)-P-P-[LIVA]-[GA]-[IV]-x(4)-[GKN] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update



		1	032
Pfam	Prosite	Full Name	Description
			July 1998 / First entry. References [1] Rudd K.E., Humphery-Smith I., Wasinger V.C., Bairoch A. Electrophoresis 19:536-544(1998).
UPF0066	PDOC01022	Uncharacterized protein family UPF0066 signature	The following uncharacterized proteins have been shown [1] to be evolutionary related: - Escherichia coli hypothetical protein yaeB and HI0510, the corresponding Haemophilus influenzae protein Agrobacterium tumefaciens Ti plasmid protein virR Pseudomonas aeruginosa protein rcsF Archaeoglobus fulgidus hypothetical protein AF0241 Archaeoglobus fulgidus hypothetical protein MJ1583 Methanococcus jannaschii hypothetical protein MJ1583 Methanobacterium thermoautotrophicum hypothetical protein MTH1797. These are proteins of from 120 to 240 amino-acid resiudes (with the exception of AF0433 which is 366 residues long). As a signature pattern we selected a conserved region in the central part of these proteins.
•			Description of pattern(s) and/or profile(s) Consensus pattern G-[AV]-F-[STA]-x-R-[SA]-x(2)-R-P-N Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / First entry. References [1] Bairoch A. Unpublished observations (1998).
UPF0076	PDOC00838	Uncharacterized protein family UPF0076 signature	The following uncharacterized proteins have been shown [1] to share regions of similarities: Goat antigen UK114, a human homolog and the rat corresponding protein which is known as perchloric acid soluble protein (PSP1). PSP1 [2] may inhibit an initiation stage of cell-free protein synthesis. Mouse heat-responsive protein HRSP12. Yeast chromosome V hypothetical protein YER057c. Yeast chromosome IX hypothetical protein YIL051c. Caenorhabditis elegans hypothetical protein C23G10.2. Escherichia coli hypothetical protein ydaK. Escherichia coli hypothetical protein yhaR. Escherichia coli hypothetical protein yigF and HI0719, the corresponding Haemophilus influenzae protein. Escherichia coli hypothetical protein yoaB. Bacillus subtilis hypothetical protein yabJ. Haemophilus influenzae hypothetical protein HI1627. Helicobacter pylori hypothetical protein HP0944. Lactococcus lactis aldR. Myxococcus xanthus dfrA. Synechocystis strain PCC 6803 hypothetical protein slr0709. Rhizobium strain NGR234 symbiotic plasmid hypothetical protein y4sK. Pyrococcus horikoshii hypothetical protein PH0854.
			highly conserved. As a signature pattern, we selected a well conserved region



		1	033
Pfam	Prosite	Full Name	Description
			located in the C-terminal part of these proteins.
			Description of pattern(s) and/or profile(s)
			Consensus pattern [PA]-[ASTPV]-R-[SACVF]-x-[LIVMFY]-x(2)- [GSAKR]-x-[LMVA]- x(5,8)-[LIVM]-E-[MI] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Last update July 1999 / Pattern and text revised. References [1] Bairoch A. Unpublished observations (1995).
			Oka T., Tsuji H., Noda C., Sakai K., Hong YM., Suzuki I., Munoz S., Natori Y. J. Biol. Chem. 270:30060-30067(1995).
UPF0099		Domain of unknown function UPF0099	Accession number: PF01981 Definition: Domain of unknown function UPF0099 Previous Pfam IDs: DUF119; Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 132.80 132.80 Noise cutoffs: -35.70 -35.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference INTERPRO; IPR002833; Comment: This domain has no known function. Number of members: 10
UQ_con	PDOC00163	Ubiquitin-conjugating enzymes active site	Ubiquitin-conjugating enzymes (EC 6.3.2.19) (UBC or E2 enzymes) [1,2,3] catalyze the covalent attachment of ubiquitin to target proteins. An activated ubiquitin moiety is transferred from an ubiquitin-activating enzyme (E1) to E2 which later ligates ubiquitin directly to substrate proteins with or without the assistance of 'N-end' recognizing proteins (E3). In most species there are many forms of UBC (at least 9 in yeast) which are implicated in diverse cellular functions. A cysteine residue is required for ubiquitin-thiolester formation. There is a single conserved cysteine in UBC's and the region around that residue is conserved in the sequence of known UBC isozymes. We have used that region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [FYWLSP]-H-[PC]-[NH]-[LIV]-x(3,4)-G-x-[LIV]-C-[LIV]-x- [LIV] [C is the active site residue] Sequences known to belong to this class detected by the pattern
			ALL, except for yeast UBC6 (DOA2). Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Jentsch S. jentsch@zmbh.uni-heidelberg.de Last update



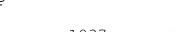
			034
Pfam	Prosite	Full Name	Description July 1998 / Text revised. References [1]
			Jentsch S., Seufert W., Sommer T., Reins HA. Trends Biochem. Sci. 15:195-198(1990).
			[2] Jentsch S., Seufert W., Hauser HP. Biochim. Biophys. Acta 1089:127-139(1991).
			[3] Hershko A. Trends Biochem. Sci. 16:265-268(1991).
urease_gamma	PDOC00133	Urease signatures	Urease (EC 3.5.1.5) is a nickel-binding enzyme that catalyzes the hydrolysis of urea to carbon dioxide and ammonia [1]. Historically, it was the first
			enzyme to be crystallized (in 1926). It is mainly found in plant seeds, microorganisms and invertebrates. In plants, urease is a hexamer
			of identical chains. In bacteria [2], it consists of either two or three different subunits (alpha, beta and gamma).
			Urease binds two nickel ions per subunit; four histidine, an aspartate and a carbamated-lysine serve as ligands to these metals; an additional histidine is
			involved in the catalytic mechanism [3].
			As signatures for this enzyme, we selected a region that contains two histidine that bind one of the nickel ions and the region of the active site histidine.
			Description of pattern(s) and/or profile(s)
			Consensus pattern T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P [The two H's bind nickel] Sequences known to belong to this class detected by the pattern ALL.
			Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-
			[LIVM]-x-F-A [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.
			Last update November 1997 / Patterns and text revised. References
			[1] Takishima K., Suga T., Mamiya G. Eur. J. Biochem. 175:151-165(1988).
			[2] Mobley H.L.T., Husinger R.P. Microbiol. Rev. 53:85-108(1989).
			[3] Jabri E., Carr M.B., Hausinger R.P., Karplus P.A. Science 268:998-1004(1995).
UreD		UreD urease accessory protein	Accession number: PF01774 Definition: UreD urease accessory protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1109 (release 4.2) Gathering cutoffs: 25 25



		1.	035
Pfam	Prosite	Full Name	Description
· tott			Trusted cutoffs: 186.00 186.00
			Noise cutoffs: -42.60 -42.60
			HMM build command line: hmmbuild -F HMM SEED
			HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 97352660
			Reference Title: Characterization of UreG, identification of a
			Reference Title: UreD-UreF-UreG complex, and evidence
			suggesting that a
			Reference Title: nucleotide-binding site in UreG is required
			for in vivo
			Reference Title: metallocenter assembly of Klebsiella
			aerogenes urease.
			Reference Author: Moncrief MB, Hausinger RP;
			Reference Location: J Bacteriol 1997;179:4081-4086.
	i		Reference Number: [2]
			Reference Medline: 96146510
			Reference Title: Organization of Ureaplasma urealyticum
			urease gene cluster
			Reference Title: and expression in a suppressor strain of
	,		Escherichia coli.
	ľ		Reference Author: Neyrolles O, Ferris S, Behbahani N,
	!		Montagnier L, Blanchard
		}	Reference Author: A;
			Reference Location: J Bacteriol 1996;178:647-655.
			Reference Number: [3]
			Reference Medline: 94211837 Reference Title: In vitro activation of urease apoprotein and
			role of UreD
			Reference Title: as a chaperone required for nickel
			metallocenter assembly.
	i		Reference Author: Park IS, Carr MB, Hausinger RP;
		1	Reference Location: Proc Natl Acad Sci U S A 1994;91:3233-
			3237.
			Database Reference INTERPRO; IPR002669;
			Comment: UreD is a urease accessory protein. Urease
			urease hydrolyses
			Comment: urea into ammonia and carbamic acid [2].
			UreD is involved in
			Comment: activation of the urease enzyme via the
			UreD-UreF-UreG-urease complex
			Comment: [1] and is required for urease nickel
			metallocenter assembly [3].
			Comment: See also UreF UreF, UreG HypB UreG.
			Number of members: 23
			5504700
UreF		UreF	Accession number: PF01730
			Definition: UreF
			Author: Bashton M, Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Pfam-B_2037 (release 4.1)
			Gathering cutoffs: -31 -31
		1	Trusted cutoffs: -14.30 -14.30
!		İ	Noise cutoffs: -49.30 -49.30
			HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
	1		
	1	1	Reference Number: [1]
]		Reference Medline: 96404789
			Reference Title: Purification and activation properties of
		1:	UreD-UreF-urease
			Reference Title: apoprotein complexes. Reference Author: Moncrief MB, Hausinger RP;
	}		
			1.14-1-1-11-1
			_
	1		urease gene cluster
			Reference Title: and expression in a suppressor strain of
			Escherichia coli.
			Reference Author: Neyrolles O, Ferris S, Behbahani N,
			Montagnier L, Blanchard
		1	Reference Author: A;
			Reference Location: J Bacteriol 1996;178:647-655.
i			Database Reference INTERPRO; IPR002639;



			1036
Pfam	Prosite	Full Name	Description The Control of the Usesse
			Comment: This family consists of the Urease accessory protein
			Comment: UreF. The urease enzyme (urea
			amidohydrolase) Comment: hydrolyses urea into ammonia and carbamic
			acid (2).
			Comment: UreF is proposed to modulate the activation
			process of Comment: urease by eliminating the binding of nickel
			irons to
			Comment: noncarbamylated protein [1]. Number of members: 20
Vif		Retroviral Vif (Viral	Accession number: PF00559 Definition: Retroviral Vif (Viral infectivity) protein
		infectivity) protein	Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Swiss-Prot Gathering cutoffs: 25 25
			Trusted cutoffs: 53.90 53.90
			Noise cutoffs: 23.60 23.60
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 95287525
			Reference Title: Aberrant Gag protein composition of a human
			Reference Title: immunodeficiency virus type 1 vif mutant
			produced in
			Reference Title: primary lymphocytes. Reference Author: Simm M, Shahabuddin M, Chao W, Allan
			JS, Volsky DJ;
			Reference Location: J Virol 1995;69:4582-4586.
			Database Reference INTERPRO; IPR000475; Comment: -!- Human immunodeficiency virus type 1
			(HIV-1) Vif is required for
			Comment: productive infection of T lymphocytes and
			macrophages. Virions Comment: produced in the absence of Vif have
			abnormal core morphology and
			Comment: those produced in primary T cells carry immature core
			proteins Comment: and low levels of mature capsid.
			Number of members: 503
Vpu		Vpu protein	Accession number: PF00558
			Definition: Vpu protein Author: Bateman A
			Alignment method of seed: Clustalw
			Source of seed members: Swiss-Prot
1			Gathering cutoffs: 15 15 Trusted cutoffs: 15.50 15.50
			Noise cutoffs: 13.60 13.60
			HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
			Reference Number: [1]
			Reference Medline: 97479365
			Reference Title: Enhancement of retroviral production from
			packaging cell Reference Title: lines expressing the human
			immunodeficiency type 1 VPU
			Reference Title: gene. Reference Author: Kobinger GP, Mouland AJ, Lalonde JP,
			Forget J. Cohen EA:
			Reference Location: Gene Ther 1997;4:868-874.
			Reference Number: [2] Reference Medline: 95156576
			Reference Title: The human immunodeficiency virus type 1
			Vou protein
			Reference Title: specifically binds to the cytoplasmic domain
			of CD4: Reference Title: implications for the mechanism of
			degradation.
			Reference Author: Bour S, Schubert U, Strebel K; Reference Location: J Virol 1995;69:1510-1520.
			Reference Location: J Virol 1995;69:1510-1520.



		1	037
Pfam	Prosite	Full Name	Description
			Reference Number: [3] Reference Medline: 97325981
			Reference Title: Secondary structure and tertiary fold of the
			human
			Reference Title: immunodeficiency virus protein U (Vpu)
			cytoptasmic domain
			Reference Title: in solution. Reference Author: Willbold D, Hoffmann S, Rosch P;
			Reference Location: Eur J Biochem 1997;245:581-588.
			Database Reference: SCOP; 1vpu; fa; [SCOP-USA][CATH-
			PDBSUM] Database Reference INTERPRO; IPR002094;
			Database Reference PDB; 1vpu; 38; 81;
			Database reference: PFAMB; PB003303;
	1		Database reference: PFAMB; PB005882;
			Comment: -!- The Vpu protein contains an N-terminal transmembrane spanning region
			Comment: and a C-terminal cytoplasmic region.
			Comment: -!- The HIV-1 Vpu protein stimulates virus
			production by enhancing
			Comment: the release of viral particles from infected cells.
			Comment: -!- The VPU protein binds specifically to
			CD4.
			Number of members: 194
XPG N	PDOC00658	XPG protein signatures	
			recessive disease,
			characterized by a high incidence of sunlight-induced skin cancer. People's
			skin cells with this condition are hypersensitive to ultraviolet
			light, due
			to defects in the incision step of DNA excision repair. There are a
			minimum of seven genetic complementation groups involved in this pathway:
			XP-A to XP-G.
			The defect in XP-G can be corrected by a 133 Kd nuclear protein
			called XPG (or XPGC) [2].
			XPG belongs to a family of proteins [2,3,4,5,6] that are
			composed of two
			main subsets:
			- Subset 1, to which belongs XPG, RAD2 from budding yeast
			and rad13 from
			fission yeast. RAD2 and XPG are single-stranded DNA
			endonucleases [7,8]. XPG makes the 3'incision in human DNA nucleotide excision
			repair [9].
		•	- Subset 2, to which belongs mouse and human FEN-1, rad2
			from fission yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific
			endonuclease.
			to the second to the second to the second to the
			In addition to the proteins listed in the above groups, this family also
			includes:
	1		Final property and the FL 21 double attended DNA overvisions
			- Fission yeast exo1, a 5'->3' double-stranded DNA exonuclease that could act
			in a pathway that corrects mismatched base pairs.
			- Yeast EXO1 (DHS1), a protein with probably the same function
			as exo1. - Yeast DIN7.
			- Teast Dilyt.
			Sequence alignment of this family of proteins reveals that
			similarities are largely confined to two regions. The first is located at the N-
			terminal
			extremity (N-region) and corresponds to the first 95 to 105 amino
			acids. The
			second region is internal (I-region) and found towards the C-terminus; it
		1	1

		1	038
Ptam	Prosite	Füll®Name	Description spans about 140 residues and contains a highly conserved core of 27 amino acids that includes a conserved pentapeptide (E-A-[DE]-A-[QS]). It is possible that the conserved acidic residues are involved in the catalytic mechanism of DNA excision repair in XPG. The amino acids linking the N- and I-regions are not conserved; indeed, they are largely absent from proteins belonging to the second subset. We have developed two signature patterns for these proteins. The first corresponds to the central part of the N-region, the second to part of the I-region and includes the putative catalytic core pentapeptide.
			Description of pattern(s) and/or profile(s) Consensus pattern [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]- [QS]-[CLM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Clarkson S.G. clarkson@medecine.unige.ch Last update November 1997 / Patterns and text revised. References [1] Tanaka K., Wood R.D.
			Trends Biochem. Sci. 19:83-86(1994). [2] Scherly D., Nouspikel T., Corlet J., Ucla C., Bairoch A., Clarkson S.G. Nature 363:182-185(1993). [3] Carr A.M., Sheldrick K.S., Murray J.M., Al-Harithy R., Watts F.Z., Lehmann A.R. Nucleic Acids Res. 21:1345-1349(1993). [4] Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S., Lehmann A.R., Carr A.M., Watts F.Z. Mol. Cell. Biol. 14:4878-4888(1994). [5] Harrington J.J., Lieber M.R.
			[6] Szankasi P., Smith G.R. Science 267:1166-1169(1995). [7] Habraken Y., Sung P., Prakash L., Prakash S. Nature 366:365-368(1993). [8] O'Donovan A., Scherly D., Clarkson S.G., Wood R.D. J. Biol. Chem. 269:15965-15968(1994).



			039
Pfam	Prosite	Full Name	Description
-			[9] O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D. Nature 371:432-435(1994):
Y_phosphatase	PDOC00323	Tyrosine specific protein phosphatases signature and profiles	Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) [1 to 5] are enzymes that catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). The currently known PTPases are listed below: Soluble PTPases. - PTPN1 (PTP-1B). PTPN2 (T. cell PTPase: TC-PTP)
			- PTPN2 (T-cell PTPase; TC-PTP) PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1- like domain (see <pdoc00566>) and could act at junctions between the membrane and cytoskeleton PTPN5 (STEP) PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes which contain two copies of the SH2 domain at its N-terminal extremity. The Drosophila protein corkscrew (gene csw) also belongs to this subgroup PTPN7 (LC-PTP; Hematopoietic protein-tyrosine phosphatase; HePTP) PTPN8 (70Z-PEP) PTPN9 (MEG2) PTPN12 (PTP-G1; PTP-P19).</pdoc00566>
			- Yeast PTP1 Yeast PTP2 which may be involved in the ubiquitin-mediated protein degradation pathway Fission yeast pyp1 and pyp2 which play a role in inhibiting the onset of mitosis Fission yeast pyp3 which contributes to the dephosphorylation of cdc2 Yeast CDC14 which may be involved in chromosome segregation Yersinia virulence plasmid PTPAses (gene yopH) Autographa californica nuclear polyhedrosis virus 19 Kd PTPase.
			Dual specificity PTPases. - DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1); which dephosphorylates MAP kinase on both Thr-183 and Tyr-185. - DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues. - DUSP3 (VHR). - DUSP3 (VHR). - DUSP4 (HVH2). - DUSP5 (HVH3). - DUSP6 (Pyst1; MKP-3). - DUSP7 (Pyst2; MKP-X). - Yeast MSG5, a PTPase that dephosphorylates MAP kinase FUS3. - Yeast YVH1. - Vaccinia virus H1 PTPase; a dual specificity phosphatase. Receptor PTPases.



Pfam	Prosite	Full Name	Description
Pfam	Prosite	Full Name	Structurally, all known receptor PTPases, are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectin type III (FN-III) repeats, immunoglobulin-like domains, MAM domains or carbonic anhydrase-like domains in their extracellular region. The cytoplasmic region generally contains two copies of the PTPAse domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not.
			PTPases is shown: Extracellular Intracellular
			Ig FN-3 CAH MAM PTPase Leukocyte common antigen (LCA) (CD45) 0 2 0 0 2 Leukocyte antigen related (LAR) 3 8 0 0 2 Drosophila DLAR 3 9 0 0 2 PTP-alpha (LRP) 0 0 0 0 2 PTP-beta 0 16 0 0 1 PTP-gamma 0 1 1 0 2 PTP-delta 0 >7 0 0 2 PTP-delta 0 >7 0 0 2 PTP-kappa 1 4 0 1 2 PTP-mu 1 4 0 1 2 PTP-mu 1 4 0 1 2 PTP-zeta 0 1 1 0 2 PTP-zeta 0 1 1 0 2 PTP-zeta 0 1 1 0 2 PTP-mu 1 1 0 1 2 PTP-zeta 0 1 1 0 2
			PTPase domains consist of about 300 amino acids. There are two conserved cysteines, the second one has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important. We derived a signature pattern for PTPase domains centered on
			the active site cysteine. There are three profiles for PTPases, the first one spans the complete domain and is not specific to any subtype. The second profile is specific to dual-specificity PTPases and the third one to the PTP subfamily.
			Description of pattern(s) and/or profile(s) Consensus pattern [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY] [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for nine sequences. Other sequence(s) detected in SWISS-PROT 3. Sequences known to belong to this class detected by the 1st profile ALL. Other sequence(s) detected in SWISS-PROT 2.



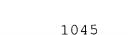
Priorite ALL dual type PTPases. Other sequence(s) detected in SWISS-PROT NONE. Sequences know to belong to this class detected by the 3rd profile ALL PTP type PTPases. Other sequence(s) detected in SWISS-PROT NONE. Note the M-phase inducer phosphatases (cdc25-type phosphatase) are tyrosine-protein phosphatases that are not structurally related to the above PTPases. Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so. Last update July 1999 / Text revised. References Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991). [2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992). [3] Trovibridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). [5] Hunter T. Cell Sei:1013-1016(1989). Zein seed storage protein Author: Batteman A Alignment method of seed: Clustalw Source of seed members: Plan-B_181 (release 4.0) Comment In the state of seed of the seed of	***************************************			D
Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the 3rd profile ALL PTP type PTPstase. Other sequence(s) detected in SWISS-PROT NONE. Note the M-phase inducer phosphatases (ode25-type phosphatase) are tyrosine- protein phosphatases that are not structurally related to the shove PTPstase. Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so. Last update July 1999 / Text revised. Put lenness 1 Flicher E.H., Charbonneau H., Tonks N.K. Science 253-401-406(1991). 1 Flicher E.H., Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8-463-493(1992). 1 3 Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). 1 4 Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14.497-500(1999). 1 5 Hunter T. Cell 55:1013-1016(1989). Accession number: PF01559 Definition: Zoin-seed storage protein Alignment method of seed: Clustalw Source of seed members: Pfam. B_181 (release 4.0) Gathering cutoffs: 4:60-46.0 Noise cutoffs: 4:60-46.0 N	Pfam	Prosite	Full Name	Description
Sequences known to belong to this class detected by the 3rd profile ALL PTP type PTPases. Other sequences is detected in SWISS-PROT NONE. Note the M-phase inducer phosphatases (cd25-type phosphatases) are syrosine-protein phosphatases that are not structurally related to the above PTPases. Note this documentation entry is linked to both a signature pattern of profiles. As profiles are much more accessive than the pattern, you should use them if you have access to the necessary software tools to do so. Last update July 1999 / Text revised. References I specifies are much more access to the necessary software tools to do so. Last update July 1999 / Text revised. References I specifies are much more access to the necessary software tools to do so. Last update July 1999 / Text revised. References I specifies a section of the sect	7000			Other sequence(s) detected in SMISS PROT NONE
profile ALL PTP type PTPases. Other sequence(s) detected in SWISS-PROT NONE. Note the M-shase inducer phosphatases (scl25 type phosphatase) are typosine-protein phosphatases) are typosine-protein phosphatases and the profiles are much more sensitive than the pattern, you should use them if you have access to the necessary but date. In the pattern of the pattern				Other sequence(s) detected in Syvido-FHOT NOINE.
profile ALL PTP type PTPases. Other sequence(s) detected in SWISS-PROT NONE. Note the M-shase inducer phosphatases (scl25 type phosphatase) are typosine-protein phosphatases) are typosine-protein phosphatases and the profiles are much more sensitive than the pattern, you should use them if you have access to the necessary but date. In the pattern of the pattern				Sequences known to belong to this class detected by the 3rd
Cither sequence(s) detected in SWISS-PROT NONE. Note the M-phase inducer phosphatases (cdc25-type phosphatase) are tyrosine- protein phosphatases that are not structurally related to the above PTPases. Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so. Last update Description of the pattern of the		!		
phosphatase) are fyrosine: protein phosphatases that are not structurally related to the above PTPases. Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary sensitive than the pattern, you should use them if you have access to the necessary sensitive than the pattern, you should use them if you have access to the necessary sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive that you have a companied to sensitive that you have a companied to sensitive than the pattern, you have a companied to sensitive that you have a companied to sensitive that you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied that you have a companied than the pattern of you have a companied than the pattern of you have a companied than the pattern of you have a companied that you have a companied th				
phosphatase) are fyrosine: protein phosphatases that are not structurally related to the above PTPases. Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary sensitive than the pattern, you should use them if you have access to the necessary sensitive than the pattern, you should use them if you have access to the necessary sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive that you have a companied to sensitive that you have a companied to sensitive than the pattern, you have a companied to sensitive that you have a companied to sensitive that you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied to sensitive than the pattern, you have a companied that you have a companied than the pattern of you have a companied than the pattern of you have a companied than the pattern of you have a companied that you have a companied th				
Structurally related to the above PTPases. Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary Last update July 1999 / Text revised. References I11 Fischer E.H., Charbonneau H., Tonks N.K. Science 253-401-406(1991). I2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol, 8:463-493(1992). I3] Trovbridge I.S. J. Biol. Chem. 269-23517-23520(1991). I4 Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1999). I5 Hunter T. Cell Sci.103-1016(1989). Zein seed storage protein Zein seed storage protein Zein seed storage protein Alignment method of seed: Clustally Source of seed members: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustally Source of seed members: PF0159 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustally Source of seed members: PF0159 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustally Source of seed members: PF0159 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustally Source of seed members: PF0159 Definition: Alignment method of seed: Clustally Source of seed members: Pfo169 Seed Source of seed members: Pfo189 Seed Source of seed members: Pf0159 Definition: Alignment method of seed: Clustally Source of seed members: Pf0168 Seed Source of seed members: SMART Glathering cutoffs: Side Source of seed members: SMART Glathering cutoffs: Side Source of seed members: SMART Glathering cutoffs: Side Source of seed members: SMART Glathering cutoffs: Side Source of seed members: SMART Glathering cutoffs: Side Source of seed members: SMART Glathering cutoffs: Side Source of seed members: SMART				
Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so. Last update Reference. Interpretation of the pattern				
and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so. Last update July 1999 / Text revised. References Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991). [2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992). [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tronks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1999). [5] Hunter T. Cell 58:1013-1016(1989). [5] Hunter T. Cell 58:1013-1016(1989). [6] Accession number PF01559 Definition: Zein seed storage protein Author: Beteman A July 1999 / Text revised cutoffs: 4:60 4:60 Author: Asserting the protein of the protein				structurally related to the above PTPases.
Last update July 1999 / Text revised. References [1] Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991). [2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:483-493(1992). [3] Trowbirdge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Scl. 14:497-500(1989). [5] Hurter T. Cell 58:1013-1016(1989). Zein seed storage protein Zein seed storage protein Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering outoffs: -21: 21 Trusted cutoffs: -40: 46.60 Noise cutoffs: -40: 46.60 Noise cutoffs: -40: 46.60 Noise cutoffs: -40: 46.60 Noise cutoffs: -40: 46.60 Noise cutoffs: -40: 46.60 HMb build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: phombuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: phombuild -F				and to profiles. As profiles are much more sensitive than the
July 1999 / Text revised. References [11] References [11] Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991). [2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992). [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). [5] Hunter T. Cell 55:1013-1016(1989). Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21:21 Trusted cutoffs: -20:40 Hill baild command line: Pfam-B_181 (release 4.0) Gathering cutoffs: -21:21 Trusted cutoffs: -40:40 Hill baild command line: Pfam-B_181 (release 4.0) Gathering cutoffs: -21:21 Trusted cutoffs: -40:40 Hill baild command line: Pfam-B_181 (release 4.0) Gathering cutoffs: -21:21 Trusted cutoffs: -40:40 Hill baild command line: Pfam-B_181 (release 4.0) Gathering cutoffs: -21:21 Trusted cutoffs: -40:40 Hill baild command line: Pfam-B_181 (release 4.0) Gathering cutoffs: -21:21 Trusted cutoffs: -21				software tools to do so.
References 11 Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991). [2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992). [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Scl. 14:497-500(1989). [5] Hunter T. Cell 58:1013-1016(1989). [5] Hunter T. Cell 58:1013-1016(1989). [5] Hunter T. Cell 58:1013-1016(1989). [6] Hunter T. Cell 58:1013-1016(1989). [7] Gathering cutoffs: 2-12. Tusted cutoffs: Alignment method of sect Clustally Source of seed members: Plams 1:81 (release 4.0) Gathering cutoffs: 4-60. 4-6.0 Noise cutoffs: 4-60. 4-6.0 Hunter Clusted Command line: hummbuild -F Hunter School -F Hunter Clusted Command line: hummbuild -F Hunter Clusted Command line: hummbuild -F Hunter Clusted Command line: hummbuild -F Hunter Clusted Command line: hummbuild -F Hunter Clusted Command line: hummbuild -F Hunter Clusted Command line: hummbuild -F Hunter Clusted Clust				·
International Content				1 -
Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991). [2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992). [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Scl. 14:497-500(1989). [5] Hunter T. Cell 58:1013-1016(1989). [5] Hunter T. Cell 58:1013-1016(1989). [6] Zein seed storage protein Author. Bateman A. Alignment method of seed: Clustally Source of seed members: Pfam.B. 181 (release 4.0) Gathering cutoffs: -21:21 Trusted cutoffs: 4.60.46.0 Noise cutoffs: 4.60.46.0 Noise cutoffs: 4.60.46.0 HMM build command line: hmmobild-F-HMM SEED HMM build command line: hmmobild-F-HMM SEED HMM build command line: hmmobild-F-HMM SEED HMM build command line: hmmobild-F-HMM SEED HMM Build command line: hmmobild-F-HMM SEED HMM build command line: hmmobild-F-HMM SEED HMM build command line: show a series of the zein-like alpha-prolamine Based on an Reference Title: Implications for their Reference Author: Arruda P; Reference Location: Database Reference Comment: Lefterence Author: Arruda P; Reference Location: Database Reference Location: Comment: Lefterence Author: Arruda P; Reference Location: Comment: Lefterence Author: Arruda P; Reference Location: Comment: Lefterence Author: Arruda P; Reference Location: Comment: Lefterence Sead members: 48 ZF-AN1 AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering ouths: SMART Gathering ouths:				
Science 253:401-406(1991). [2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992). [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). [5] Hunter T. Cell 58:1013-1016(1989). Zein seed storage protein Author: Bateman A. Alignment method of seed: Clustalw Source of seed members: Plam-B_181 (release 4.0) Gathering cutofis: -21-21 Trusted cutoffs: -4.60 -4.60 HiMM build command line: Immibuild -F HMM SEED HiMM build command line: Immibuild -F HMM SEED HiMM build command line: Immibuild -F HMM SEED HiMM build command line: Immibuild -F HMM SEED HiMM Build command line: Immibuild -F HMM				
Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 6:463-493(1992). [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). [5] Hunter T. Cell 58:1013-1016(1989). Zein seed storage protein Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustallw Source of seed members: Plame-B_181 (release 4.0) Gathering cutoffs: -21-21 Trusted cutoffs: -46.06 -46.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.06 -46.00 Noise cutoffs: -46.06 -46.00 Noise cutoffs: -46.06 -46.00 Source of seed members: Plame-B_181 (release 4.0) Gathering cutoffs: -46.06 -46.00 Noise				
Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 6:463-493(1992). [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). [5] Hunter T. Cell 58:1013-1016(1989). Zein seed storage protein Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustallw Source of seed members: Plame-B_181 (release 4.0) Gathering cutoffs: -21-21 Trusted cutoffs: -46.06 -46.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.04-6.00 Noise cutoffs: -46.06 -46.00 Noise cutoffs: -46.06 -46.00 Noise cutoffs: -46.06 -46.00 Source of seed members: Plame-B_181 (release 4.0) Gathering cutoffs: -46.06 -46.00 Noise				l
Annu. Rev. Cell Biol. 8:463-493(1992). [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). [5] Hunter T. Cell 58:1013-1016(1989). Zein seed storage protein Author: Bateman A. Alignment method of seed Sustalw Source of seed members: 'PF01559 Definition: Author: Bateman A. Alignment method of seed: Clustalw Source of seed members: 'PanB-B_181 (release 4.0) Gathering cutoffs: '20: 21: 21 Trusted cutoffs: '40: 40: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '46: 60: 46: 60 Noise cutoffs: '40: 40: 40: 40: 40: 40: 40: 40: 40: 40:				
I 3 Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). I 4 Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). I 5 Hunter T. Cell 58:1013-1016(1989). Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: 4.60 4.60 Noise cutoffs: 4.60 4.60 Noise cutoffs: 4.60 4.60 Noise cutoffs: 4.60 4.60 HMM build command line: Immedibrateseed 0 HMM Reference Number: I Reference Medilen: 93197294 Reference Title: Implications for their Reference Title: Implications for their Reference Title: Reference Title: Reference Title: Reference Title: Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: sequences show a series of tandem repeats II Number of members: 48 AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Alignment method of seed: Manual Source of seed members: 16 16 16 16 16 16 16 16				
Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). [5] Hunter T. Cell 58:1013-1016(1989). Zein seed storage protein Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Plam-B_181 (release 4.0) Gathering cutoffs: -21:-21 Trusted cutoffs: -46:-60 Noise cutoffs: -46:-60 Noise cutoffs: -46:-60 Noise cutoffs: -46:-60 Noise cutoffs: -46:-60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM seed on an anatysis of amino and the cheen command line: hmmbuild -F HMM seed on an anatys				74111d. 1107. Golf Biol. G. 166 (1662).
Zein seed storage protein Zein seed storage protein Zein seed storage protein Zein seed storage protein Zein seed storage protein Zein seed storage protein Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PFam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -4.60 -46.60 HMM build command line: hmmouild -F HMM SEED HMM build command line: hmmouild -F HMM				
L4 Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). [5]				
Zein seed storage protein Zein seed storage protein Zein seed storage protein Zein seed storage protein Zein seed storage protein Author: Bateman A Alignment method of seed: Clustallw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -4.12 -1.21 Trusted cutoffs: -4.60 -4.60 Noise cutoffs: -4.60	1			J. Biol. Chem. 266:23517-23520(1991).
Zein seed storage protein Zein seed storage protein Zein seed storage protein Zein seed storage protein Zein seed storage protein Author: Bateman A Alignment method of seed: Clustallw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -4.12 -1.21 Trusted cutoffs: -4.60 -4.60 Noise cutoffs: -4.60				[4]
Zein seed storage protein Zein seed storage protein Zein seed storage protein Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PF01428 Definition: AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Arruda P; Accession number: PF01428 Definition: AN1-like Zinc finger Author: AN1-like Zinc finger Author: AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 1616				
Zein Seed storage protein Zein seed storage protein Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.0 -46.60 -46.60 HMM build command line: hmmobuild -F HMM SEED HMM build command line: hmmobuild -F HMM SEED HMM build command line: hmmobuild -F HMM SEED HMM build command line: hmmobuild -F HMM SEED HMM build command line: hmmobuild -F HMM SEED HMM build command line: hmmobuild -F HMM SEED HMM build command line: hmmobuild -F HMM SEED HMM build command line: hmmobuild -F HMM SEED HMM build command line: hmmobuild -F HMM SEED HMM build command line: hmmobuild -F HMM SEED HMM Beference Number: [1] 93197294 Reference Medline: Studies of the zein-like alpha-prolamins based on an Reference Title: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16				· · · · · · · · · · · · · · · · · · ·
Zein seed storage protein Zein seed storage protein Zein seed storage protein Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21-21 Trusted cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: based on an Reference Title: studies of the zein-like alpha-prolamins based on an Reference Title: evolution and three-dimensional structure. Arruda P; Reference Location: Journal of the color of the protein seed storage proteins. They are unusually rich in Comment: residues and their Comment: sequences show a series of tandem repeats [1] AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16				
Zein Seed storage protein Zein seed storage protein Zein seed storage protein Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.60 -46.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM build command line: hmmbuild -F HMM build command line: hmmbuild -F HMM build command line: hmmbuild -F HMM build command line: hmmbuild -F HMM build command line: hmmbuild -F HMM build command line: hmmbuild -F HMM build command line: hmmbuild -F HMM build command line: hmmbuild -F HMM build -F HMM build -F HMM build -F HMM build command line: hmmbuild -F HMM buil				
Zein seed storage protein Zein seed storage protein Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfan-B_181 (release 4.0) Gathering cutoffs: -21 - 21 Trusted cutoffs: -46.60 - 46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Mediline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: Implications for their Reference Author: Arruda P; Reference Author: Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Comment: condition of the comment: comment: residues and their Comment: glutamine, proline, alanine, and leucine sequences show a series of tandem repeats [11] AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: MART Alignment method of seed: Manual Source of seed members: 16 16				
protein Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.0 4.60 Noise cutoffs: -46.0 -46.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM S				Cell 36:1013-1016(1989).
protein Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.0 4.60 Noise cutoffs: -46.0 -46.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM S	Zoin		Zein seed storage	Accession number: PF01559
Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.60 -46.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM SEED HMM Self-Seed I HMM Stude of the zein-like alpha-prolamins sudvition and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P, Proteins 1993;15:88-99. INTERPRO; IPRO02530; Zeins are seed storage proteins. They are unusually rich in Comment: unusually rich in Comment: sequences show a series of tandem repeats [1] Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16	Zelli	1	Zoni seeu siviage	
Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -4.60 -4.6.0 Noise cutoffs: -4.6.0 -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: Reference Number: [1] Reference Medline: 93197294 Studies of the zein-like alpha-prolamins based on an Reference Title: malysis of amino acid sequences: implications for their Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: glutamine, proline, alanine, and leucine sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PFQ1428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: MaRT Gathering cutoffs: 16 16	1	l .	protein	Definition: Zein seed storage protein
Gathering cutoffs: -21 - 21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 - 46.60 HMM build command line: hmmbuild - F HMM SEED HMM build command line: hmmbuild - F HMM Seed 0 HMM Studies of the zein-like alpha-prolamins analysis of amino acid sequences: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Lotation: Database Reference Comment: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Lite: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Lite: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Lite: evolution and three-dim			protein	
Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: based on an Reference Title: implications for their Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger AN1-like Zinc finger AN1-like Zinc finger AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw
Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild -F HMM Seed 0 HMM [1] 93197294 Studies of the zein-like alpha-prolamins analysis of amino acid sequences: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Proteins 1993;15:88-99. INTERPRO; IPRO02530; INTERPRO; IPRO0			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0)
HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate seed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: based on an Reference Title: implications for their Reference Title: analysis of amino acid sequences: implications for their Reference Title: reference Title: analysis of amino acid sequences: implications for their Reference Title: reference Coation: Proteins 1993;15:88-99. Database Reference Comment: Louisually rich in Comment: residues and their Comment: residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21
HMM build command line: hmmcalibrateseed 0 HMM Reference Number: Reference Medline: Reference Title: based on an Reference Title: implications for their Reference Title: Reference Title: analysis of amino acid sequences: implications for their Reference Title: Reference Title: Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: residues and their Comment: 11]. Number of members: 48 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Accession number: Summan AN1-like Zinc finger Accession number: Summan AN1-like Zinc finger Accession number: Summan AN1-like Zinc finger Accession number: Summan AN1-like Zinc finger Accession number: Summan AN1-like Zinc finger Accession number: Summan AN1-like Zinc finger Accession number: Summan AN1-like Zinc finger Accession number: Summan AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60
Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: implications for their Reference Title: Reference Title: Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. AN1-like Zinc finger Reference Number: Studies of the zein-like alpha-prolamins based on an Reference Title: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Proteins 1993;15:88-99. INTERPRO; IPR002530; Zeins are seed storage proteins. They are glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60
Reference Title: based on an Reference Title: implications for their Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: Zf-AN1 AN1-like Zinc finger Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: glutamine, proline, alanine, and leucine sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED
based on an Reference Title: implications for their Reference Title: evolution and three-dimensional structure. Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: residues and their Comment: [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM
Reference Title: implications for their Reference Title: Reference Title: Reference Title: Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger AN1-like Zinc finger AN1-like Zinc finger AN1-like Zinc finger AN1-like Zinc finger ACCESSION NUMBER: PF0,1428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294
implications for their Reference Title: Reference Author: Arruda P; Reference Location: Database Reference Comment: Unusually rich in Comment: residues and their Comment: Sequences show a series of tandem repeats [1]. Number of members: AN1-like Zinc finger Accession number: Definition: Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 1 evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Proteins 1993;15:88-99. INTERPRO; IPR002530; Zeins are seed storage proteins. They are glutamine, proline, alanine, and leucine residues and their Sequences show a series of tandem repeats [1]. Number of members: 48 AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins
Reference Title: Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: [1]. Number of members: AN1-like Zinc finger AN1-like Zinc finger Reference Author: Reference Author: Reference Author: Arruda P; Reference Location: Database Reference Location: INTERPRO; IPR002530; Zeins are seed storage proteins. They are unusually rich in Comment: residues and their Comment: Sequences show a series of tandem repeats [1]. Number of members: 48 AN1-like Zinc finger Accession number: Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an
Reference Author: Arruda P; Reference Location: Database Reference Comment: Unusually rich in Comment: residues and their Comment: [1]. Number of members: AN1-like Zinc finger AN1-like Zinc finger Reference Author: Arruda P; Reference Location: Proteins 1993;15:88-99. INTERPRO; IPR002530; Zeins are seed storage proteins. They are unusually rich in Comment: residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 AN1-like Zinc finger Accession number: Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences:
Reference Location: Proteins 1993;15:88-99. Database Reference Comment: Zeins are seed storage proteins. They are unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their
Database Reference INTERPRO; IPR002530; Comment: Zeins are seed storage proteins. They are unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Reference Author: Garratt R, Oliva G, Caracelli I, Leite A,
Comment: Zeins are seed storage proteins. They are unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF0,1428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Reference Author: Garratt R, Oliva G, Caracelli I, Leite A,
unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1], Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: salpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Reference Author: Arruda P; Reference Location: Proteins 1993;15:88-99.
Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Author: Arruda P; Reference Location: Proteins 1993;15:88-99. INTERPRO; IPR002530;
Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Zf-AN1 AN1-like Zinc finger Accession number: PF0,1428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: Proteins 1993;15:88-99. INTERPRO; IPR002530; Zeins are seed storage proteins. They are
zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Nedline: 93197294 Reference Title: 93197294 Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: INTERPRO; IPR002530; Zeins are seed storage proteins. They are
Zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: 93197294 Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: Unusually rich in Comment: glutamine, proline, alanine, and leucine
zf-AN1 AN1-like Zinc finger Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: [1] Reference Title: 93197294 Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Author: Arruda P; Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats
Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16			protein	Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: [1] Reference Title: 93197294 Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Author: Arruda P; Reference Author: Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1].
Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16				Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.60 -46.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: [1] Reference Medline: 93197294 Reference Title: 93197294 Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Author: Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: residues and their Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48
Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16	zf-AN1			Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: based on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: Unusually rich in Comment: residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Accession number: PF01428
Source of seed members: SMART Gathering cutoffs: 16 16	zf-AN1			Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: [1] Reference Title: 93197294 Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Author: Arruda P; Reference Author: Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Accession number: PF01428 Definition: AN1-like Zinc finger
Gathering cutoffs: 16 16	zf-AN1			Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: [1] Reference Title: 93197294 Reference Title: analysis of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Author: Arruda P; Reference Author: Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART
Trusted cutoffs: 16.40 16.40	zf-AN1			Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Number: [1] Reference Title: 93197294 Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Author: Arruda P; Reference Author: Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual
	zf-AN1			Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: -46.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Nedline: 93197294 Reference Medline: 93197294 Reference Title: shased on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Database Reference Comment: comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48 Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16

	Procito	Full Name	Description
fam	Prosite		Noise cutoffs: 7.30 7.30 Holian command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] 93292985 Reference Title: Two related localized mRNAs from Kenopus laevis encode Reference Title: ubiquitin-like fusion proteins. Linnen JM, Bailey CP, Weeks DL; Gene 1993;128:181-188. SMART; ZnF_AN1; INTERPRO; IPR000058; Zinc finger at the C-terminus of An1 Swiss:Q91889, a ubiquitin-like Comment: protein in Xenopus laevis. Comment: The following pattern describes the zinc finger. Comment: C-X2-C-X(9-12)-C-X(1-2)-C-X4-C-X2-H-X5-H-X-C Comment: indicate the number of residues. Number of members: 18
zf-B_box	PDOC50015	B-box zinc finger	Accession number: PF00643 Definition: B-box zinc finger. Author: Bateman A Alignment method of seed: pftools Source of seed members: Prosite Gathering cutoffs: 25 25 Trusted cutoffs: 26.00 26.00 Noise cutoffs: 24.50 29.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Database Reference: SCOP; 1fre; fa; [SCOP-USA][CATH-PBSUM] Database Reference: PROSITE_PROFILE; PS50119; Database Reference: PROSITE; PDOC50015 Database Reference: Database Reference: PDB; 1fre; 4; 42; Database reference: PFAMB; PB002777; Database reference: PFAMB; PB010625; Database reference: Number of members: 44
zf-CONSTANS		CONSTANS family zinc finger	Accession number: PF01760 Definition: CONSTANS family zinc finger Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1072 (release 4.2) Gathering cutoffs: 25 10 Trusted cutoffs: 76.10 17.20 Noise cutoffs: 9.70 9.70 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 95211836 Reference Title: The CONSTANS gene of Arabidopsis promotes flowering and Reference Title: encodes a protein showing similarities to zinc finger Reference Author: Putterill J, Robson F, Lee K, Simon R, Coupland G; Reference Location: Database Reference Number of members: 45
zf-DHHC		DHHC zinc finger domain	n Accession number: PF01529 Definition: DHHC zinc finger domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_945 (release 4.0) Gathering cutoffs: 22 22



		1	043	
Pfam	Prosite	Full Name	Description	
				22.40 22.40
				-22.40 -22.40
				line: hmmbuild HMM SEED
				line: hmmcalibrateseed 0 HMM
			Reference Number:	[1]
			Reference Medline:	99250263
			Reference Title:	The drosophila STAM gene homolog is in a
			tight gene	about a sud the events are correlated to that
			Reference Title:	cluster, and its expression correlates to that
			of the	adjacent gene ial.
			Reference Title: Reference Author:	Mesilaty-Gross S, Reich A, Motro B,
			Wides R:	Mesilary-Gloss S, Melch A, Morro B,
	•		Reference Location:	Gene 1999;231:173-186.
			Reference Number:	[2]
			Reference Medline:	97315340
			Reference Title:	Variations of the C2H2 zinc finger motif in
			the yeast	
			Reference Title:	genome and classification of yeast zinc
			finger proteins.	general and an an an an an an an an an an an an an
			Reference Author:	Bohm S, Frishman D, Mewes HW;
			Reference Location:	Nucleic Acids Res 1997;25:2464-2469.
	}		Reference Number:	[3]
			Reference Medline:	99321009
			Reference Title:	The DHHC domain: a new highly conserved
			cysteine-rich	
			Reference Title:	motif.
			Reference Author:	Putilina T, Wong P, Gentleman S;
		ļ	Reference Location:	Mol Cell Biochem 1999;195:219-226.
			Reference Number:	[4]
			Reference Medline:	10490616
			Reference Title:	Erf2, a Novel Gene Product That Affects the
			Localization	
			Reference Title:	and Palmitoylation of Ras2 in
			Saccharomyces cere	
		†	Reference Author:	Bartels DJ, Mitchell DA, Dong X,
			Deschenes RJ;	
			Reference Location:	Mol Cell Biol 1999;19:6775-6787.
			Database Reference	INTERPRO; IPR001594;
•			Comment:	This domain is also known as NEW1 [2].
			This domain is	The
			Comment:	predicted to be a zinc binding domain. The
			function	of this domain is unknown, but it has been
			Comment:	of this domain is unknown, but it has been
			predicted to	be involved in protein-protein or protein-DNA
			Comment:	interactions [3].
·			Number of members:	• •
			Number of members.	. 54
- LAVAID		MYND finger	Accession number:	PF01753
zf-MYND		WITHD IIIIger		1YND finger
				ateman A
			Alignment method of	
			Source of seed mem	
			Gathering cutoffs:	11 11
			Trusted cutoffs:	17.30 17.30
1			Noise cutoffs:	5.50 5.50
			HMM build command	line: hmmbuild HMM SEED
		İ		line: hmmcalibrateseed 0 HMM
		1	Reference Number:	[1]
			Reference Medline:	96203118
			Reference Title:	DEAF-1, a novel protein that binds an
1			essential region in a	
			Reference Title:	Deformed response element.
	1		Reference Author:	Gross CT, McGinnis W;
			Reference Location:	
			Reference Number:	[2]
		1	Reference Medline:	98079069
		1	Reference Title:	Molecular cloning, sequence analysis,
			expression, and	Alance distribution of accompanies a married
	1		Reference Title:	tissue distribution of suppressin, a novel
			suppressor of	cell cycle entry
1			Reference Title: Reference Author:	cell cycle entry. LeBoeuf RD, Ban EM, Green MM, Stone

		Ψ,	0 4 4
Pfam	Prosite	Full Name	Description
			AS, Propst SM, Blalock Reference Author: JE, Tauber JD; Reference Location: J Biol Chem 1998;273:361-368. Database Reference INTERPRO; IPR002893; Number of members: 48
Zn_carbOpept	PDOC00123	Zinc carboxypeptidases, zinc-binding regions signatures	There are a number of different types of zinc-dependent carboxypeptidases (EC 3.4.17) [1,2]. All these enzymes seem to be structurally and functionally related. The enzymes that belong to this family are listed below. - Carboxypeptidase A1 (EC 3.4.17.1), a pancreatic digestive enzyme that can removes all C-terminal amino acids with the exception of Arg, Lys and Pro Carboxypeptidase A2 (EC 3.4.17.15), a pancreatic digestive enzyme with a specificity similar to that of carboxypeptidase A1, but with a preference for bulkier C-terminal residues Carboxypeptidase B (EC 3.4.17.2), also a pancreatic digestive enzyme, but that preferentially removes C-terminal Arg and Lys Carboxypeptidase N (EC 3.4.17.3) (also known as arginine carboxypeptidase), a plasma enzyme which protects the body from potent vasoactive and inflammatory peptides containing C-terminal Arg or Lys (such as kinins or anaphylatoxins) which are released into the circulation Carboxypeptidase B, an enzyme located in secretory granules of pancreatic islets, adrenal gland, pituitary and brain. This enzyme removes residual C-terminal Arg or Lys remaining after initial endoprotease cleavage during prohormone processing Carboxypeptidase M (EC 3.4.17.12), a membrane bound Arg and Lys specific enzyme. It is ideally situated to act on peptide hormones at local tissue sites where it could control their activity before or after interaction with specificity plasma membrane receptors Mast cell carboxypeptidase A, but found in the secretory granules of mast cells Streptomyces griseus carboxypeptidase (Cpase SG) (EC 3.4.17.1), an enzyme with a specificity to carboxypeptidase A but found in the secretory granules of mast cells Streptomyces griseus carboxypeptidase (Cpase SG) (EC 3.4.17.1) (CPT) [4], which also combines the specificities of mammalian carboxypeptidases A and B Thermoactinomyces vulgaris carboxypeptidase T (EC 3.4.17.18) (CPT) [4], which also combines the specificities of carboxypeptidases A and B AEBP1 (Sp. a transcriptiona



		1	045
Yam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s)
			Consensus pattern [PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]- [STAG]-x(6)-[LIVMFYTA] [H and E are zinc ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT Bacillus sphaericus endopeptidase I which hydrolyses the gamma-D-Glu-(L)meso-diaminopimelic acid bond of spore cortex peptidoglycan [6] and which is possibly distantly related to zinc carboxypeptidases.
			Consensus pattern H-[STAG]-x(3)-[LIVME]-x(2)-[LIVMFYW]-P- [FYW] [H is a zinc ligand] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 40.
			Note if a protein includes both signatures, the probability of it being a eukaryotic zinc carboxypeptidase is 100%
			Note these proteins belong to families M14A/M14B in the classification of peptidases [7,E1]. Last update November 1995 / Patterns and text revised. References
			[1] Tan F., Chan S.J., Steiner D.F., Schilling J.W., Skidgel R.A. J. Biol. Chem. 264:13165-13170(1989).
			[2] Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F., Serafin W.E. J. Biol. Chem. 264:20094-20099(1989).
			[3] Narahashi Y. J. Biochem. 107:879-886(1990).
			[4] Teplyakov A., Polyakov K., Obmolova G., Strokopytov B., Kuranova I., Osterman A.L., Grishin N.V., Smulevitch S.V., Zagnitko O.P., Galperina O.V., Matz M.V., Stepanov V.M. Eur. J. Biochem. 208:281-288(1992).
			[5] He GP., Muise A., Li A.W., Ro HS. Nature 378:92-96(1995).
			[6] Hourdou ML., Guinand M., Vacheron M.J., Michel G., Denoroy L., Duez C.M., Englebert S., Joris B., Weber G., Ghuysen JM. Biochem. J. 292:563-570(1993).
			[7] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).
			[E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt
ZZ		Zinc finger present in dystrophin, CBP/p300	Accession number: PF00569 Definition: Zinc finger present in dystrophin, CBP/p300 Author: SMART Alignment method of seed: Manual Source of seed members: Alignment kindly provided by SMART Gathering cutoffs: 14 14 Trusted cutoffs: 14.60 14.60 Noise cutoffs: 10.90 10.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrateseed 0 HMM Reference Number: [1] Reference Medline: 96402609
			Reference Medline: 96402609 Reference Title: ZZ and TAZ: new putative zinc fingers in



Pfam	Prosite	Full Name Description
		dystrophin and Reference Title: other proteins. Reference Author: Ponting CP, Blake DJ, Davies KE, Kendrick-Jones J, Winder Reference Author: SJ; Reference Location: Trends Biochem Sci 1996;21:11-13.
		Database Reference: EXPERT; Chris.Ponting@human- anatomy.oxford.ac.uk; Database Reference: INTERPRO; IPR000433; Database reference: PFAMB; PB041629; Comment: ZZ in dystrophin binds calmodulin
		Comment: Putative zinc finger; binding not yet showr Number of members: 87

10

15

20

25

30

Attorney No. 2750-1237P 9/689,980 PP 1049-1078-endoggia

AA. Activities of Polypeptides Comprising Signal Peptides

Polypeptides comprising signal peptides are a family of proteins that are typically targeted to (1) a particular organelle or intracellular compartment, (2) interact with a particular molecule or (3) for secretion outside of a host cell. Example of polypeptides comprising signal peptides include, without limitation, secreted proteins, soluble proteins, receptors, proteins retained in the ER, etc.

These proteins comprising signal peptides are useful to modulate ligand-receptor interactions, cell-to-cell communication, signal transduction, intracellular communication, and activities and/or chemical cascades that take part in an organism outside or within of any particular cell.

One class of such proteins are soluble proteins which are transported out of the cell. These proteins can act as ligands that bind to receptor to trigger signal transduction or to permit communication between cells.

Another class is receptor proteins which also comprise a retention domain that lodges the receptor protein in the membrane when the cell transports the receptor to the surface of the cell. Like the soluble ligands, receptors can also modulate signal transduction and communication between cells.

In addition the signal peptide itself can serve as a ligand for some receptors. An example is the interaction of the ER targeting signal peptide with the signal recognition particle (SRP). Here, the SRP binds to the signal peptide, halting translation, and the resulting SRP complex then binds to docking proteins located on the surface of the ER, prompting transfer of the protein into the ER.

A description of signal peptide residue composition is described below in Subsection IV.C.1.

5

15

20

25

30

III. **Methods of Modulating Polypeptide Production**

It is contemplated that polynucleotides of the invention can be incorporated into a host cell or in-vitro system to modulate polypeptide production. For instance, the SDFs prepared as described herein can be used to prepare expression cassettes useful in a number of techniques for suppressing or enhancing expression.

An example are polynucleotides comprising sequences to be transcribed, such as coding sequences, of the present invention can be inserted into nucleic acid constructs to modulate polypeptide production. Typically, such sequences to be transcribed are heterologous to at least one element of the nucleic acid construct to generate a chimeric gene or construct.

Another example of useful polynucleotides are nucleic acid molecules comprising regulatory sequences of the present invention. Chimeric genes or constructs can be generated when the regulatory sequences of the invention linked to heterologous sequences in a vector construct. Within the scope of invention are such chimeric gene and/or constructs.

Also within the scope of the invention are nucleic acid molecules, whereof at least a part or fragment of these DNA molecules are presented in Tables 1 and 2 of the present application, and wherein the coding sequence is under the control of its own promoter and/or its own regulatory elements. Such molecules are useful for transforming the genome of a host cell or an organism regenerated from said host cell for modulating polypeptide production.

Additionally, a vector capable of producing the oligonucleotide can be inserted into the host cell to deliver the oligonucleotide.

More detailed description of components to be included in vector constructs are described both above and below.

Whether the chimeric vectors or native nucleic acids are utilized, such polynucleotides can be incorporated into a host cell to modulate polypeptide production. Native genes and/or nucleic acid molecules can be effective when exogenous to the host cell.

Methods of modulating polypeptide expression includes, without limitation: Suppression methods, such as

Antisense

Ribozymes

Co-suppression

Insertion of Sequences into the Gene to be Modulated

Regulatory Sequence Modulation.

10

15

20

25

as well as Methods for Enhancing Production, such as
Insertion of Exogenous Sequences; and
Regulatory Sequence Modulation.

III.A. Suppression

Expression cassettes of the invention can be used to suppress expression of endogenous genes which comprise the SDF sequence. Inhibiting expression can be useful, for instance, to tailor the ripening characteristics of a fruit (Oeller et al., *Science* 254:437 (1991)) or to influence seed size_(WO98/07842) or to provoke cell ablation (Mariani et al., Nature 357: 384-387 (1992).

As described above, a number of methods can be used to inhibit gene expression in plants, such as antisense, ribozyme, introduction of exogenous genes into a host cell, insertion of a polynucleotide sequence into the coding sequence and/or the promoter of the endogenous gene of interest, and the like.

III.A.1. Antisense

An expression cassette as described above can be transformed into host cell or plant to produce an antisense strand of RNA. For plant cells, antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, *see*, e.g., Sheehy et al., *Proc. Nat. Acad. Sci.* USA, 85:8805 (1988), and Hiatt et al., U.S. Patent No. 4,801,340.

III.A.2. Ribozymes

Similarly, ribozyme constructs can be transformed into a plant to cleave mRNA and down-regulate translation.

III.A.3. Co-Suppression

Another method of suppression is by introducing an exogenous copy of the gene to be suppressed. Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter has been shown to prevent the accumulation of mRNA. A detailed description of this method is described above.

III.A.4. <u>Insertion of Sequences into the Gene to be Modulated</u>

10

15

20

25

30

Yet another means of suppressing gene expression is to insert a polynucleotide into the gene of interest to disrupt transcription or translation of the gene.

Homologous recombination could be used to target a polynucleotide insert to a gene using the Cre-Lox system (A.C. Vergunst et al., *Nucleic Acids Res.* 26:2729 (1998), A.C. Vergunst et al., *Plant Mol. Biol.* 38:393 (1998), H. Albert et al., *Plant J.* 7:649 (1995)).

In addition, random insertion of polynucleotides into a host cell genome can also be used to disrupt the gene of interest. Azpiroz-Leehan et al., *Trends in Genetics* 13:152 (1997). In this method, screening for clones from a library containing random insertions is preferred for identifying those that have polynucleotides inserted into the gene of interest. Such screening can be performed using probes and/or primers described above based on sequences from Tables 1 and 2, fragments thereof, and substantially similar sequence thereto. The screening can also be performed by selecting clones or any transgenic plants having a desired phenotype.

III.A.5. Regulatory Sequence Modulation

The SDFs described in Tables 1 and 2, and fragments thereof are examples of nucleotides of the invention that contain regulatory sequences that can be used to suppress or inactivate transcription and/or translation from a gene of interest as discussed in I.C.5.

III.A.6. Genes Comprising Dominant-Negative Mutations

When suppression of production of the endogenous, native protein is desired it is often helpful to express a gene comprising a dominant negative mutation. Production of protein variants produced from genes comprising dominant negative mutations is a useful tool for research Genes comprising dominant negative mutations can produce a variant polypeptide which is capable of competing with the native polypeptide, but which does not produce the native result. Consequently, over expression of genes comprising these mutations can titrate out an undesired activity of the native protein. For example, The product from a gene comprising a dominant negative mutation of a receptor can be used to constitutively activate or suppress a signal transduction cascade, allowing examination of the phenotype and thus the trait(s) controlled by that receptor and pathway. Alternatively, the protein arising from the gene comprising a dominant-negative mutation can be an inactive enzyme still capable of binding to the same substrate as the native protein and therefore competes with such native protein.

Products from genes comprising dominant-negative mutations can also act upon the native protein itself to prevent activity. For example, the native protein may be active only as a homo-multimer or as one subunit of a hetero-multimer. Incorporation of an inactive subunit into the multimer with native subunit(s) can inhibit activity.

Thus, gene function can be modulated in host cells of interest by insertion into these cells vector constructs comprising a gene comprising a dominant-negative mutation.

III.B. Enhanced Expression

Enhanced expression of a gene of interest in a host cell can be accomplished by either (1) insertion of an exogenous gene; or (2) promoter modulation.

III.B.1. Insertion of an Exogenous Gene

Insertion of an expression construct encoding an exogenous gene can boost the number of gene copies expressed in a host cell.

Such expression constructs can comprise genes that either encode the native protein that is of interest or that encode a variant that exhibits enhanced activity as compared to the native protein. Such genes encoding proteins of interest can be constructed from the sequences from Tables 1 and 2, fragments thereof, and substantially similar sequence thereto.

Such an exogenous gene can include either a constitutive promoter permitting expression in any cell in a host organism or a promoter that directs transcription only in particular cells or times during a host cell life cycle or in response to environmental stimuli.

III.B.2. Regulatory Sequence Modulation

The SDFs of Tables 1 and 2, and fragments thereof, contain regulatory sequences that can be used to enhance expression of a gene of interest. For example, some of these sequences contain useful enhancer elements. In some cases, duplication of enhancer elements or insertion of exogenous enhancer elements will increase expression of a desired gene from a particular promoter. As other examples, all ll promoters require binding of a regulatory protein to be activated, while some promoters may need a protein that signals a promoter binding protein to expose a polymerase binding site. In either case, over-production of such proteins can be used to enhance expression of a gene of interest by increasing the activation time of the promoter.

Such regulatory proteins are encoded by some of the sequences in Tables 1 and 2, fragments thereof, and substantially similar sequences thereto.

10

5

15

20

30

10

15

20

25

30

Coding sequences for these proteins can be constructed as described above.

IV. Gene Constructs and Vector Construction

To use isolated SDFs of the present invention or a combination of them or parts and/or mutants and/or fusions of said SDFs in the above techniques, recombinant DNA vectors which comprise said SDFs and are suitable for transformation of cells, such as plant cells, are usually prepared. The SDF construct can be made using standard recombinant DNA techniques (Sambrook et al. 1989) and can be introduced to the species of interest by *Agrobacterium*-mediated transformation or by other means of transformation (*e.g.*, particle gun bombardment) as referenced below.

The vector backbone can be any of those typical in the art such as plasmids, viruses, artificial chromosomes, BACs, YACs and PACs and vectors of the sort described by

- (a) **BAC:** Shizuya et al., Proc. Natl. Acad. Sci. USA 89: 8794-8797 (1992); Hamilton et al., Proc. Natl. Acad. Sci. USA 93: 9975-9979 (1996);
 - (b) YAC: Burke et al., Science 236:806-812 (1987);.
 - (c) PAC: Sternberg N. et al., Proc Natl Acad Sci U S A. Jan;87(1):103-7 (1990);
- (d) **Bacteria-Yeast Shuttle Vectors:** Bradshaw et al., Nucl Acids Res 23: 4850-4856 (1995);
- (e) Lambda Phage Vectors: Replacement Vector, e.g.,
 Frischauf et al., J. Mol Biol 170: 827-842 (1983); or Insertion vector, e.g.,
 Huynh et al., In: Glover NM (ed) DNA Cloning: A practical Approach, Vol.1 Oxford: IRL
 Press (1985);
 - (f) **T-DNA gene fusion vectors :** Walden et al., Mol Cell Biol 1: 175-194 (1990); and
 - (g) Plasmid vectors: Sambrook et al., infra.

Typically, a vector will comprise the exogenous gene, which in its turn comprises an SDF of the present invention to be introduced into the genome of a host cell, and which gene may be an antisense construct, a ribozyme construct chimeraplast, or a coding sequence with any desired transcriptional and/or translational regulatory sequences, such as promoters, UTRs, and 3' end termination sequences. Vectors of the invention can also include origins of replication, scaffold attachment regions (SARs), markers, homologous sequences, introns, etc.

A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational

10

15

20

25

30

initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for over-expression, a plant promoter fragment may be employed that will direct transcription of the gene in all tissues of a regenerated plant. Alternatively, the plant promoter may direct transcription of an SDF of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters).

If proper polypeptide production is desired, a polyadenylation region at the 3'-end of the coding region is typically included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences from genes or SDF or the invention may comprise a marker gene that confers a selectable phenotype on plant cells. The vector can include promoter and coding sequence, for instance. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or phosphinotricin.

IV.A. Coding Sequences

Generally, the sequence in the transformation vector and to be introduced into the genome of the host cell does not need to be absolutely identical to an SDF of the present invention. Also, it is not necessary for it to be full length, relative to either the primary transcription product or fully processed mRNA. Furthermore, the introduced sequence need not have the same intron or exon pattern as a native gene. Also, heterologous non-coding segments can be incorporated into the coding sequence without changing the desired amino acid sequence of the polypeptide to be produced.

IV.B. Promoters

As explained above, introducing an exogenous SDF from the same species or an orthologous SDF from another species can modulate the expression of a native gene corresponding to that SDF of interest. Such an SDF construct can be under the control of either a constitutive promoter or a highly regulated inducible promoter (e.g., a copper inducible promoter). The promoter of interest can initially be either endogenous or heterologous to the species in question. When re-introduced into the genome of said species, such promoter becomes exogenous to said species. Over-expression of an SDF transgene can

10

15

20

25

30

lead to co-suppression of the homologous endogeneous sequence thereby creating some alterations in the phenotypes of the transformed species as demonstrated by similar analysis of the chalcone synthase gene (Napoli et al., *Plant Cell* 2:279 (1990) and van der Krol et al., *Plant Cell* 2:291 (1990)). If an SDF is found to encode a protein with desirable characteristics, its over-production can be controlled so that its accumulation can be manipulated in an organ- or tissue-specific manner utilizing a promoter having such specificity.

Likewise, if the promoter of an SDF (or an SDF that includes a promoter) is found to be tissue-specific or developmentally regulated, such a promoter can be utilized to drive or facilitate the transcription of a specific gene of interest (e.g., seed storage protein or root-specific protein). Thus, the level of accumulation of a particular protein can be manipulated or its spatial localization in an organ- or tissue-specific manner can be altered.

IV. C Signal Peptides

SDFs of the present invention containing signal peptides are indicated in Tables 1 and 2. In some cases it may be desirable for the protein encoded by an introduced exogenous or orthologous SDF to be targeted (1) to a particular organelle intracellular compartment, (2) to interact with a particular molecule such as a membrane molecule or (3) for secretion outside of the cell harboring the introduced SDF. This will be accomplished using a signal peptide.

Signal peptides direct protein targeting, are involved in ligand-receptor interactions and act in cell to cell communication. Many proteins, especially soluble proteins, contain a signal peptide that targets the protein to one of several different intracellular compartments. In plants, these compartments include, but are not limited to, the endoplasmic reticulum (ER), mitochondria, plastids (such as chloroplasts), the vacuole, the Golgi apparatus, protein storage vessicles (PSV) and, in general, membranes. Some signal peptide sequences are conserved, such as the Asn-Pro-Ile-Arg amino acid motif found in the N-terminal propeptide signal that targets proteins to the vacuole (Marty (1999) *The Plant Cell* 11: 587-599). Other signal peptides do not have a consensus sequence *per se*, but are largely composed of hydrophobic amino acids, such as those signal peptides targeting proteins to the ER (Vitale and Denecke (1999) *The Plant Cell* 11: 615-628). Still others do not appear to contain either a consensus sequence or an identified common secondary sequence, for instance the chloroplast stromal targeting signal peptides (Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). Furthermore, some targeting peptides are bipartite, directing proteins first to an organelle and then to a membrane within the organelle (e.g. within the thylakoid lumen of the

20

25

5

10



chloroplast; see Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). In addition to the diversity in sequence and secondary structure, placement of the signal peptide is also varied. Proteins destined for the vacuole, for example, have targeting signal peptides found at the N-terminus, at the C-terminus and at a surface location in mature, folded proteins. Signal peptides also serve as ligands for some receptors.

These characteristics of signal proteins can be used to more tightly control the phenotypic expression of introduced SDFs. In particular, associating the appropriate signal sequence with a specific SDF can allow sequestering of the protein in specific organelles (plastids, as an example), secretion outside of the cell, targeting interaction with particular receptors, etc. Hence, the inclusion of signal proteins in constructs involving the SDFs of the invention increases the range of manipulation of SDF phenotypic expression. The nucleotide sequence of the signal peptide can be isolated from characterized genes using common molecular biological techniques or can be synthesized in vitro.

In addition, the native signal peptide sequences, both amino acid and nucleotide, described in Tables 1 and 2 can be used to modulate polypeptide transport. Further variants of the native signal peptides described in Tables 1 and 2 are contemplated. Insertions, deletions, or substitutions can be made. Such variants will retain at least one of the functions of the native signal peptide as well as exhibiting some degree of sequence identity to the native sequence.

Also, fragments of the signal peptides of the invention are useful and can be fused with other signal peptides of interest to modulate transport of a polypeptide.

V. Transformation Techniques

A wide range of techniques for inserting exogenous polynucleotides are known for a number of host cells, including, without limitation, bacterial, yeast, mammalian, insect and plant cells.

Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. *See, e.g.* Weising et al., *Ann. Rev. Genet.* 22:421 (1988); and Christou, Euphytica, v. 85, n.1-3:13-27, (1995).

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and

10

15

20

25

30

introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria (McCormac et al., *Mol. Biotechnol.* 8:199 (1997); Hamilton, *Gene* 200:107 (1997)); Salomon et al. *EMBO J.* 2:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983).

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *EMBO J.* 3:2717 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:773 (1987). *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary or cointegrate vectors, are well described in the scientific literature. See, for example Hamilton, *CM.*, *Gene* 200:107 (1997); Müller et al. *Mol. Gen. Genet.* 207:171 (1987); Komari et al. *Plant J.* 10:165 (1996); Venkateswarlu et al. *Biotechnology* 9:1103 (1991) and Gleave, *AP.*, *Plant Mol. Biol.* 20:1203 (1992); Graves and Goldman, *Plant Mol. Biol.* 7:34 (1986) and Gould et al., *Plant Physiology* 95:426 (1991).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as seedlessness. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture* in Handbook of Plant Cell Culture," pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1988. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467 (1987). Regeneration of monocots (rice) is described by Hosoyama et al. (*Biosci. Biotechnol. Biochem.* 58:1500 (1994)) and by Ghosh et al. (*J. Biotechnol.* 32:1 (1994)). The nucleic acids of the invention can be used to confer desired traits on essentially any plant.

Thus, the invention has use over a broad range of plants, including species from the genera Anacardium, Arachis, Asparagus, Atropa, Avena, Brassica, Citrus, Citrullus, Capsicum, Carthamus, Cocos, Coffea, Cucumis, Cucurbita, Daucus, Elaeis, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoscyamus, Lactuca, Linum, Lolium, Lupinus,

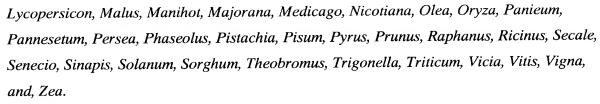
10

15

20

25

30



One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

The particular sequences of SDFs identified are provided in the attached Tables 1 and 2. One of ordinary skill in the art, having this data, can obtain cloned DNA fragments, synthetic DNA fragments or polypeptides constituting desired sequences by recombinant methodology known in the art or described herein.

EXAMPLES

The invention is illustrated by way of the following examples. The invention is not limited by these examples as the scope of the invention is defined solely by the claims following.

EXAMPLE 1: cDNA PREPARATION

A number of the nucleotide sequences disclosed in Tables 1 and 2 herein as representative of the SDFs of the invention can be obtained by sequencing genomic DNA (gDNA) and/or cDNA from corn plants grown from HYBRID SEED # 35A19, purchased from Pioneer Hi-Bred International, Inc., Supply Management, P.O. Box 256, Johnston, Iowa 50131-0256.

A number of the nucleotide sequences disclosed in Tables 1 and 2 herein as representative of the SDFs of the invention can also be obtained by sequencing genomic DNA from *Arabidopsis thaliana*, Wassilewskija ecotype or by sequencing cDNA obtained from mRNA from such plants as described below. This is a true breeding strain. Seeds of the plant are available from the Arabidopsis Biological Resource Center at the Ohio State University, under the accession number CS2360. Seeds of this plant were deposited under the terms and conditions of the Budapest Treaty at the American Type Culture Collection, Manassas, VA on August 31, 1999, and were assigned ATCC No. PTA-595.

Other methods for cloning full-length cDNA are described, for example, by Seki et al., *Plant Journal* 15:707-720 (1998) High-efficiency cloning of Arabidopsis full-length

10

15

20

25

30

cDNA by biotinylated Cap trapper"; Maruyama et al., *Gene* 138:171 (1994) Oligo-capping a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides"; and WO 96/34981.

Tissues were, or each organ was, individually pulverized and frozen in liquid nitrogen. Next, the samples were homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed. Then the sample was applied to a 2M sucrose cushion to isolate polysomes. The RNA was isolated by treatment with detergents and proteinase K followed by ethanol precipitation and centrifugation. The polysomal RNA from the different tissues was pooled according to the following mass ratios: 15/15/1 for male inflorescences, female inflorescences and root, respectively. The pooled material was then used for cDNA synthesis by the methods described below.

Starting material for cDNA synthesis for the exemplary corn cDNA clones with sequences presented in Tables 1 and 2 was poly(A)-containing polysomal mRNAs from inflorescences and root tissues of corn plants grown from HYBRID SEED # 35A19. Male inflorescences and female (pre-and post-fertilization) inflorescences were isolated at various stages of development. Selection for poly(A) containing polysomal RNA was done using oligo d(T) cellulose columns, as described by Cox and Goldberg, Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The quality and the integrity of the polyA+ RNAs were evaluated.

Starting material for cDNA synthesis for the exemplary *Arabidopsis* cDNA clones with sequences presented in Tables 1 and 2 was polysomal RNA isolated from the top-most inflorescence tissues of *Arabidopsis thaliana* Wassilewskija (Ws.) and from roots of *Arabidopsis thaliana* Landsberg erecta (L. er.), also obtained from the Arabidopsis Biological Resource Center. Nine parts inflorescence to every part root was used, as measured by wet mass. Tissue was pulverized and exposed to liquid nitrogen. Next, the sample was homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed and the sample was applied to a 2M sucrose cushion to isolate polysomal RNA. Cox et al., Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The polysomal RNA was used



10

15 .

20

25

for cDNA synthesis by the methods described below. Polysomal mRNA was then isolated as described above for corn cDNA. The quality of the RNA was assessed electrophoretically.

Following preparation of the mRNAs from various tissues as described above, selection of mRNA with intact 5' ends and specific attachment of an oligonucleotide tag to the 5' end of such mRNA was performed using either a chemical or enzymatic approach. Both techniques take advantage of the presence of the cap" structure, which characterizes the 5' end of most intact mRNAs and which comprises a guanosine generally methylated once, at the 7 position.

The chemical modification approach involves the optional elimination of the 2', 3'-cis diol of the 3' terminal ribose, the oxidation of the 2', 3'-cis diol of the ribose linked to the cap of the 5' ends of the mRNAs into a dialdehyde, and the coupling of the such obtained dialdehyde to a derivatized oligonucleotide tag. Further detail regarding the chemical approaches for obtaining mRNAs having intact 5' ends are disclosed in International Application No. WO96/34981 published November 7, 1996.

The enzymatic approach for ligating the oligonucleotide tag to the intact 5' ends of mRNAs involves the removal of the phosphate groups present on the 5' ends of uncapped incomplete mRNAs, the subsequent decapping of mRNAs having intact 5' ends and the ligation of the phosphate present at the 5' end of the decapped mRNA to an oligonucleotide tag. Further detail regarding the enzymatic approaches for obtaining mRNAs having intact 5' ends are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultés et perspectives nouvelles. Apports pour l'étude de la régulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EPO 625572 and Kato et al., Gene 150:243-250 (1994).

In both the chemical and the enzymatic approach, the oligonucleotide tag has a restriction enzyme site (e.g. an EcoRI site) therein to facilitate later cloning procedures. Following attachment of the oligonucleotide tag to the mRNA, the integrity of the mRNA is examined by performing a Northern blot using a probe complementary to the oligonucleotide tag.

10

15

20

25

30

For the mRNAs joined to oligonucleotide tags using either the chemical or the enzymatic method, first strand cDNA synthesis is performed using an oligo-dT primer with reverse transcriptase. This oligo-dT primer can contain an internal tag of at least 4 nucleotides, which can be different from one mRNA preparation to another. Methylated dCTP is used for cDNA first strand synthesis to protect the internal EcoRI sites from digestion during subsequent steps. The first strand cDNA is precipitated using isopropanol after removal of RNA by alkaline hydrolysis to eliminate residual primers.

Second strand cDNA synthesis is conducted using a DNA polymerase, such as Klenow fragment and a primer corresponding to the 5' end of the ligated oligonucleotide. The primer is typically 20-25 bases in length. Methylated dCTP is used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following second strand synthesis, the full-length cDNAs are cloned into a phagemid vector, such as pBlueScript™ (Stratagene). The ends of the full-length cDNAs are blunted with T4 DNA polymerase (Biolabs) and the cDNA is digested with EcoRI. Since methylated dCTP is used during cDNA synthesis, the EcoRI site present in the tag is the only hemi-methylated site; hence the only site susceptible to EcoRI digestion. In some instances, to facilitate subcloning, an Hind III adapter is added to the 3' end of full-length cDNAs.

The full-length cDNAs are then size fractionated using either exclusion chromatography (AcA, Biosepra) or electrophoretic separation which yields 3 to 6 different fractions. The full-length cDNAs are then directionally cloned either into pBlueScript[™] using either the EcoRI and SmaI restriction sites or, when the Hind III adapter is present in the full-length cDNAs, the EcoRI and Hind III restriction sites. The ligation mixture is transformed, preferably by electroporation, into bacteria, which are then propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached to full-length cDNAs are selected as follows.

The plasmid cDNA libraries made as described above are purified (e.g. by a column available from Qiagen). A positive selection of the tagged clones is performed as follows. Briefly, in this selection procedure, the plasmid DNA is converted to single stranded DNA using phage F1 gene II endonuclease in combination with an exonuclease (Chang et al., *Gene* 127:95 (1993)) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA is then purified using paramagnetic beads as described by Fry et al., *Biotechniques* 13: 124 (1992). Here the single stranded DNA is hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide tag. Preferably, the primer has a length of 20-25 bases. Clones including a sequence complementary to the biotinylated

10

15

oligonucleotide are selected by incubation with streptavidin coated magnetic beads followed by magnetic capture. After capture of the positive clones, the plasmid DNA is released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as ThermoSequenase™ (obtained from Amersham Pharmacia Biotech). Alternatively, protocols such as the Gene Trapper™ kit (Gibco BRL) can be used. The double stranded DNA is then transformed, preferably by electroporation, into bacteria. The percentage of positive clones having the 5' tag oligonucleotide is typically estimated to be between 90 and 98% from dot blot analysis.

Following transformation, the libraries are ordered in microtiter plates and sequenced. The *Arabidopsis* library was deposited at the American Type Culture Collection on January 7, 2000 as *E-coli* liba 010600" under the accession number **PTA-1161**.

EXAMPLE 2: SOUTHERN HYBRIDIZATIONS

The SDFs of the invention can be used in Southern hybridizations as described above. The following describes extraction of DNA from nuclei of plant cells, digestion of the nuclear DNA and separation by length, transfer of the separated fragments to membranes, preparation of probes for hybridization, hybridization and detection of the hybridized probe.

The procedures described herein can be used to isolate related polynucleotides or for diagnostic purposes. Moderate stringency hybridization conditions, as defined above, are described in the present example. These conditions result in detection of hybridization between sequences having at least 70% sequence identity. As described above, the hybridization and wash conditions can be changed to reflect the desired percenatge of sequence identity between probe and target sequences that can be detected.

In the following procedure, a probe for hybridization is produced from two PCR reactions using two primers from genomic sequence of *Arabidopsis thaliana*. As described above, the particular template for generating the probe can be any desired template.

The first PCR product is assessed to validate the size of the primer to assure it is of the expected size. Then the product of the first PCR is used as a template, with the same pair of primers used in the first PCR, in a second PCR that produces a labeled product used as the probe.

Fragments detected by hybridization, or other bands of interest, can be isolated from gels used to separate genomic DNA fragments by known methods for further purification and/or characterization.

25

30

Buffers for nuclear DNA extraction

1. 10X HB

	1000 ml	
40 mM spermidine	10.2 g	Spermine (Sigma S-2876) and spermidine (Sigma S-2501)
10 mM spermine	3.5 g	Stabilize chromatin and the nuclear membrane
0.1 M EDTA (disodium)	37.2 g	EDTA inhibits nuclease
0.1 M Tris	12.1 g	Buffer
0.8 M KCl	59.6 g	Adjusts ionic strength for stability of nuclei

Adjust pH to 9.5 with 10 N NaOH. It appears that there is a nuclease present in leaves. Use of pH 9.5 appears to inactivate this nuclease.

2. 2 M sucrose (684 g per 1000 ml)

Heat about half the final volume of water to about 50°C. Add the sucrose slowly then bring the mixture to close to final volume; stir constantly until it has dissolved. Bring the solution to volume.

3. Sarkosyl solution (lyses nuclear membranes)

Adjust the pH to 9.5 after all the components are dissolved and bring up to the proper volume.

4. 20% Triton X-100

80 ml Triton X-100

320 ml 1xHB (w/o β -ME and PMSF)

Prepare in advance; Triton takes some time to dissolve

5 A. Procedure

1. Prepare 1X H" buffer (keep ice-cold during use)

1000 ml

10X HB 100 ml

2 M sucrose 250 ml a non-ionic osmoticum

Water 634 ml

Added just before use:

100 mM PMSF* 10 ml a protease inhibitor; protects

nuclear membrane proteins

ß-mercaptoethanol 1 ml inactivates nuclease by reducing

disulfide bonds

*100 mM PMSF

(phenyl methyl sulfonyl fluoride, Sigma P-7626) (add 0.0875 g to 5 ml 100% ethanol)

- 2. Homogenize the tissue in a blender (use 300-400 ml of 1xHB per blender). Be sure that you use 5-10 ml of HB buffer per gram of tissue. Blenders generate heat so be sure to keep the homogenate cold. It is necessary to put the blenders in ice periodically.
- 3. Add the 20% Triton X-100 (25 ml per liter of homogenate) and gently stir on ice for 20 min. This lyses plastid, but not nuclear, membranes.

10

15

COETTO, JOHESTO

5

10

15

20

- 4. Filter the tissue suspension through several nylon filters into an ice-cold beaker. The first filtration is through a 250-micron membrane; the second is through an 85-micron membrane; the third is through a 50-micron membrane; and the fourth is through a 20-micron membrane. Use a large funnel to hold the filters. Filtration can be sped up by gently squeezing the liquid through the filters.
- 5. Centrifuge the filtrate at 1200 x g for 20 min. at 4°C to pellet the nuclei.
- 6. Discard the dark green supernatant. The pellet will have several layers to it. One is starch; it is white and gritty. The nuclei are gray and soft. In the early steps, there may be a dark green and somewhat viscous layer of chloroplasts.

Wash the pellets in about 25 ml cold H buffer (with Triton X-100) and resuspend by swirling gently and pipetting. After the pellets are resuspended.

Pellet the nuclei again at 1200 - 1300 x g. Discard the supernatant.

Repeat the wash 3-4 times until the supernatant has changed from a dark green to a pale green. This usually happens after 3 or 4 resuspensions. At this point, the pellet is typically grayish white and very slippery. The Triton X-100 in these repeated steps helps to destroy the chloroplasts and mitochondria that contaminate the prep.

Resuspend the nuclei for a final time in a total of 15 ml of H buffer and transfer the suspension to a sterile 125 ml Erlenmeyer flask.

- 7. Add 15 ml, dropwise, cold 2% Sarkosyl, 0.1 M Tris, 0.04 M EDTA solution (pH 9.5) while swirling gently. This lyses the nuclei. The solution will become very viscous.
- 8. Add 30 grams of CsCl and gently swirl at room temperature until the CsCl is in solution. The mixture will be gray, white and viscous.
- 9. Centrifuge the solution at 11,400 x g at 4°C for at least 30 min. The longer this spin is, the firmer the protein pellicle.

1.0

- 10. The result is typically a clear green supernatant over a white pellet, and (perhaps) under a protein pellicle. Carefully remove the solution under the protein pellicle and above the pellet. Determine the density of the solution by weighing 1 ml of solution and add CsCl if necessary to bring to 1.57 g/ml. The solution contains dissolved solids (sucrose etc) and the refractive index alone will not be an accurate guide to CsCl concentration.
- 11. Add 20 µl of 10 mg/ml EtBr per ml of solution.
- 12. Centrifuge at 184,000 x g for 16 to 20 hours in a fixed-angle rotor.
- 13. Remove the dark red supernatant that is at the top of the tube with a plastic transfer pipette and discard. Carefully remove the DNA band with another transfer pipette.

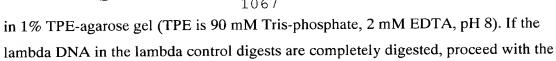
 The DNA band is usually visible in room light; otherwise, use a long wave UV light to locate the band.
- 14. Extract the ethidium bromide with isopropanol saturated with water and salt. Once the solution is clear, extract at least two more times to ensure that all of the EtBr is gone. Be very gentle, as it is very easy to shear the DNA at this step. This extraction may take a while because the DNA solution tends to be very viscous. If the solution is too viscous, dilute it with TE.
- Dialyze the DNA for at least two days against several changes (at least three times) of TE (10 mM Tris, 1mM EDTA, pH 8) to remove the cesium chloride.
 - 16. Remove the dialyzed DNA from the tubing. If the dialyzed DNA solution contains a lot of debris, centrifuge the DNA solution at least at 2500 x g for 10 min. and carefully transfer the clear supernatant to a new tube. Read the A260 concentration of the DNA.
- 25 17. Assess the quality of the DNA by agarose gel electrophoresis (1% agarose gel) of the DNA. Load 50 ng and 100 ng (based on the OD reading) and compare it with known

and good quality DNA. Undigested lambda DNA and a lambda-HindIII-digested DNA are good molecular weight makers.

Protocol for Digestion of Genomic DNA

Protocol:

- 1. The relative amounts of DNA for different crop plants that provide approximately a balanced number of genome equivalent is given in Table 3. Note that due to the size of the wheat genome, wheat DNA will be underrepresented. Lambda DNA provides a useful control for complete digestion.
- 2. Precipitate the DNA by adding 3 volumes of 100% ethanol. Incubate at -20°C for at least two hours. Yeast DNA can be purchased and made up at the necessary concentration, therefore no precipitation is necessary for yeast DNA.
- 3. Centrifuge the solution at 11,400 x g for 20 min. Decant the ethanol carefully (be careful not to disturb the pellet). Be sure that the residual ethanol is completely removed either by vacuum desiccation or by carefully wiping the sides of the tubes with a clean tissue.
- 4. Resuspend the pellet in an appropriate volume of water. Be sure the pellet is fully resuspended before proceeding to the next step. This may take about 30 min.
- 5. Add the appropriate volume of 10X reaction buffer provided by the manufacturer of the restriction enzyme to the resuspended DNA followed by the appropriate volume of enzymes. Be sure to mix it properly by slowly swirling the tubes.
- 20 6. Set-up the lambda digestion-control for each DNA that you are digesting.
 - 7. Incubate both the experimental and lambda digests overnight at 37°C. Spin down condensation in a microfuge before proceeding.
 - 8. After digestion, add 2 μl of loading dye (typically 0.25% bromophenol blue, 0.25% xylene cyanol in 15% Ficoll or 30% glycerol) to the lambda-control digests and load



precipitation of the genomic DNA in the digests.

Precipitate the digested DNA by adding 3 volumes of 100% ethanol and incubating in 9. -20°C for at least 2 hours (preferably overnight).

EXCEPTION: Arabidopsis and yeast DNA are digested in an appropriate volume; they don't have to be precipitated.

Resuspend the DNA in an appropriate volume of TE (e.g., $22 \mu l \times 50 \text{ blots} = 1100 \mu l$) 10. and an appropriate volume of 10X loading dye (e.g., $2.4 \mu l \times 50 \text{ blots} = 120 \mu l$). Be careful in pipetting the loading dye - it is viscous. Be sure you are pipetting the correct volume.

Table 3 Some guide points in digesting genomic DNA.

	Genome	Size Relative to Arabidopsis	Genome Equivalent to 2 μg Arabidopsis DNA	Amount of DNA per blot
Species	Size			
Arabidopsis	120 Mb	1X	1X	2 μg
Brassica	1,100 Mb	9.2X	0.54X	10 μg
Corn	2,800 Mb	23.3X	0.43X	20 μg
Cotton	2,300 Mb	19.2X	0.52X	20 μg
Oat	11,300 Mb	94X	0.11X	20 μg
Rice	400 Mb	3.3X	0.75X	5 μg
Soybean	1,100 Mb	9.2X	0.54X	10 μg
Sugarbeet	758 Mb	6.3X	0.8X	10 μg
Sweetclover	1,100 Mb	9.2X	0.54X	10 μg
Wheat	16,000 Mb	133X	0.08X	20 μg

10

Yeast	15 Mb	0.12X	1X	0.25 μg

Protocol for Southern Blot Analysis

The digested DNA samples are electrophoresed in 1% agarose gels in 1x TPE buffer. Low voltage; overnight separations are preferred. The gels are stained with EtBr and photographed.

- 1. For blotting the gels, first incubate the gel in 0.25 N HCl (with gentle shaking) for about 15 min.
- 2. Then briefly rinse with water. The DNA is denatured by 2 incubations. Incubate (with shaking) in 0.5 M NaOH in 1.5 M NaCl for 15 min.
- 3. The gel is then briefly rinsed in water and neutralized by incubating twice (with shaking) in 1.5 M Tris pH 7.5 in 1.5 M NaCl for 15 min.
- 4. A nylon membrane is prepared by soaking it in water for at least 5 min, then in 6X SSC for at least 15 min. before use. (20x SSC is 175.3 g NaCl, 88.2 g sodium citrate per liter, adjusted to pH 7.0.)
- The nylon membrane is placed on top of the gel and all bubbles in between are removed. The DNA is blotted from the gel to the membrane using an absorbent medium, such as paper toweling and 6x SCC buffer. After the transfer, the membrane may be lightly brushed with a gloved hand to remove any agarose sticking to the surface.
- 6. The DNA is then fixed to the membrane by UV crosslinking and baking at 80°C. The membrane is stored at 4°C until use.
- B. Protocol for PCR Amplification of Genomic Fragments in Arabidopsis

Amplification procedures:

1. Mix the following in a 0.20 ml PCR tube or 96-well PCR plate:

5

10

20

15



	1069	Final Amount or Conc.
		Final Amount of Conc.
Volume	Stock	
0.5 μl	~ 10 ng/μl genomic DNA ¹	5 ng
2.5 μl	10X PCR buffer	20 mM Tris, 50 mM KCl
0.75 μl	50 mM MgCl ₂	1.5 mM
1 μl	10 pmol/μl Primer 1 (Forward)	10 pmol
1 µl	10 pmol/μl Primer 2 (Reverse)	10 pmol
0.5 µl	5 mM dNTPs	0.1 mM
0.1 μl	5 units/µl Platinum Taq™ (Life Technologies, Gaithersburg, MD)	1 units
	DNA Polymerase	
(to 25 μl)	Water	

- 2. The template DNA is amplified using a Perkin Elmer 9700 PCR machine:
- 1) 94°C for 10 min. followed by

2)	3)	4)
5 cycles:	5 cycles:	25 cycles:
94 °C - 30 sec	94 °C - 30 sec	94 °C - 30 sec
62 °C - 30 sec	58 °C - 30 sec	53 °C - 30 sec
72 °C - 3 min	72 °C - 3 min	72 °C - 3 min

¹ Arabidopsis DNA is used in the present experiment, but the procedure is a general one.

cesses.lolzco

5

10

. 15

20

5) 72°C for 7 min. Then the reactions are stopped by chilling to 4°C.

The procedure can be adapted to a multi-well format if necessary.

Quantification and Dilution of PCR Products:

- 1. The product of the PCR is analyzed by electrophoresis in a 1% agarose gel. A linearized plasmid DNA can be used as a quantification standard (usually at 50, 100, 200, and 400 ng). These will be used as references to approximate the amount of PCR products. HindIII-digested Lambda DNA is useful as a molecular weight marker. The gel can be run fairly quickly; e.g., at 100 volts. The standard gel is examined to determine that the size of the PCR products is consistent with the expected size and if there are significant extra bands or smeary products in the PCR reactions.
- 2. The amounts of PCR products can be estimated on the basis of the plasmid standard.
- 3. For the small number of reactions that produce extraneous bands, a small amount of DNA from bands with the correct size can be isolated by dipping a sterile 10-µl tip into the band while viewing though a UV Transilluminator. The small amount of agarose gel (with the DNA fragment) is used in the labeling reaction.

C. Protocol for PCR-DIG-Labeling of DNA

Solutions:

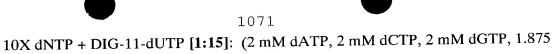
Reagents in PCR reactions (diluted PCR products, 10X PCR Buffer, 50 mM MgCl₂, 5 U/ μ l Platinum Taq Polymerase, and the primers)

10X dNTP + DIG-11-dUTP [1:5]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.65 mM dTTP, 0.35 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP [1:10]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.81 mM dTTP, 0.19 mM DIG-11-dUTP)

10

15



TE buffer (10 mM Tris, 1 mM EDTA, pH 8)

mM dTTP, 0.125 mM DIG-11-dUTP)

Maleate buffer: In 700 ml of deionized distilled water, dissolve 11.61 g maleic acid and 8.77 g NaCl. Add NaOH to adjust the pH to 7.5. Bring the volume to 1 L. Stir for 15 min. and sterilize.

10% blocking solution: In 80 ml deionized distilled water, dissolve 1.16g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, Cat. no. 1096176). Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

1% blocking solution: Dilute the 10% stock to 1% using the maleate buffer.

Buffer 3 (100 mM Tris, 100 mM NaCl, 50 mM MgCl₂, pH9.5). Prepared from autoclaved solutions of 1M Tris pH 9.5, 5 M NaCl, and 1 M MgCl₂ in autoclaved distilled water.

1. PCR reactions are performed in 25 μl volumes containing:

PCR buffer

1X

MgCl₂

1.5 mM

10X dNTP + DIG-11-dUTP

1X (please see the note below)

Platinum Taq™ Polymerase

1 unit

10 pg probe DNA

10 pmol primer 1

Note:

Use for:

10X dNTP + DIG-11-dUTP (1:5)

< 1 kb

10X dNTP + DIG-11-dUTP (1:10)

1 kb to 1.8 kb

10X dNTP + DIG-11-dUTP (1:15)

> 1.8 kb

- 2. The PCR reaction uses the following amplification cycles:
 - 1) 94°C for 10 min.

2)	3)	4)
5 cycles:	5 cycles:	25 cycles:
95°C - 30 sec	95°C - 30 sec	95°C - 30 sec
61°C - 1 min	59°C - 1 min	51°C - 1 min
73°C - 5 min	75°C - 5 min	73°C - 5 min

- 5) 72°C for 8 min. The reactions are terminated by chilling to 4°C (hold).
- 3. The products are analyzed by electrophoresis- in a 1% agarose gel, comparing to an aliquot of the unlabelled probe starting material.
- 4. The amount of DIG-labeled probe is determined as follows:

5

10

OSMOSANO. HOLEOG



Make serial dilutions of the diluted control DNA in dilution buffer (TE: 10 mM Tris and 1 mM EDTA, pH 8) as shown in the following table:

DIG-labeled control DNA starting conc.		Final Conc. (Dilution
	Stepwise Dilution	Name)
5 ng/μl	1 μl in 49 μl TE	100 pg/μl (A)
100 pg/μl (A)	25 μl in 25 μl TE	50 pg/μl (B)
50 pg/μl (B)	25 μl in 25 μl TE	25 pg/μl (C)
25 pg/μl (C)	20 μl in 30 μl TE	10 pg/μl (D)

- a. Serial deletions of a DIG-labeled standard DNA ranging from 100 pg to 10 pg are spotted onto a positively charged nylon membrane, marking the membrane lightly with a pencil to identify each dilution.
- b. Serial dilutions (e.g., 1:50, 1:2500, 1:10,000) of the newly labeled DNA probe are spotted.
- c. The membrane is fixed by UV crosslinking.
- d. The membrane is wetted with a small amount of maleate buffer and then incubated in 1% blocking solution for 15 min at room temp.
- e. The labeled DNA is then detected using alkaline phosphatase conjugated anti-DIG antibody (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) and an NBT substrate according to the manufacture's instruction.
- f. Spot intensities of the control and experimental dilutions are then compared to estimate the concentration of the PCR-DIG-labeled probe.

5

D. Prehybridization and Hybridization of Southern Blots

Solutions:

100% Formamide

purchased from Gibco

20X SSC

 $(1X = 0.15 \text{ M NaCl}, 0.015 \text{ M Na}_3\text{citrate})$

per L:

175 g NaCl

87.5 g Na₃citrate 2H ₂0

20% Sarkosyl (N-lauroyl-sarcosine)

20% SDS (sodium dodecyl sulphate)

10% Blocking Reagent: In 80 ml deionized distilled water, dissolve 1.16 g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder. Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

Prehybridization Mix:

Final		Volume	
Concentration	Components	(per 100 ml)	Stock
50%	Formamide	50 ml	100%
5X	SSC	25 ml	20X
0.1%	Sarkosyl	0.5 ml	20%
0.02%	SDS	0.1 ml	20%
2%	Blocking Reagent	20 ml	10%
	Water	4.4 ml	

General Procedures:

1. Place the blot in a heat-sealable plastic bag and add an appropriate volume of prehybridization solution (30 ml/100cm²) at room temperature. Seal the bag with a heat sealer, avoiding bubbles as much as possible. Lay down the bags in a large plastic tray (one tray can accommodate at least 4–5 bags). Ensure that the bags are

5

lying flat in the tray so that the prehybridization solution is evenly distributed throughout the bag. Incubate the blot for at least 2 hours with gentle agitation using a waver shaker.

- 2. Denature DIG-labeled DNA probe by incubating for 10 min. at 98°C using the PCR machine and immediately cool it to 4°C.
 - 3. Add probe to prehybridization solution (25 ng/ml; 30 ml = 750 ng total probe) and mix well but avoid foaming. Bubbles may lead to background.
 - 4. Pour off the prehybridization solution from the hybridization bags and add new prehybridization and probe solution mixture to the bags containing the membrane.
 - 5. Incubate with gentle agitation for at least 16 hours.
 - 6. Proceed to medium stringency post-hybridization wash:

Three times for 20 min. each with gentle agitation using 1X SSC, 1% SDS at 60°C.

All wash solutions must be prewarmed to 60°C. Use about 100 ml of wash solution per membrane.

To avoid background keep the membranes fully submerged to avoid drying in spots; agitate sufficiently to avoid having membranes stick to one another.

7. After the wash, proceed to immunological detection and CSPD development.

E. Procedure for Immunological Detection with CSPD Solutions:

20

Buffer 1:

Maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl;

adjusted to pH 7.5 with NaoH)

Washing buffer:

Maleic acid buffer with 0.3% (v/v) Tween 20.

10

15

Blocking stock solution

10% blocking reagent in buffer 1. Dissolve (10X concentration): blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, cat. no. 1096176) by constantly stirring on a 65°C heating block or heat in a microwave, autoclave and store at 4°C.

Buffer 2

(1X blocking solution):

Dilute the stock solution 1:10 in Buffer 1.

Detection buffer:

0.1 M Tris, 0.1 M NaCl, pH 9.5

Procedure:

1. After the post-hybridization wash the blots are briefly rinsed (1-5 min.) in the maleate washing buffer with gentle shaking.

2. Then the membranes are incubated for 30 min. in Buffer 2 with gentle shaking.

3. Anti-DIG-AP conjugate (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) at 75 mU/ml (1:10,000) in Buffer 2 is used for detection. 75 ml of solution can be used for 3 blots.

- 4. The membrane is incubated for 30 min. in the antibody solution with gentle shaking.
- 5. The membrane are washed twice in washing buffer with gentle shaking. About 250 mls is used per wash for 3 blots.
- 6. The blots are equilibrated for 2-5 min in 60 ml detection buffer.
- Dilute CSPD (1:200) in detection buffer. (This can be prepared ahead of time and stored in the dark at 4°C).

The following steps must be done individually. Bags (one for detection and one for exposure) are generally cut and ready before doing the following steps.

10

15

20

25

- 8. The blot is carefully removed from the detection buffer and excess liquid removed without drying the membrane. The blot is immediately placed in a bag and 1.5 ml of CSPD solution is added. The CSPD solution can be spread over the membrane. Bubbles present at the edge and on the surface of the blot are typically removed by gentle rubbing. The membrane is incubated for 5 min. in CSPD solution.
- 9. Excess liquid is removed and the membrane is blotted briefly (DNA side up) on Whatman 3MM paper. Do not let the membrane dry completely.
- 10. Seal the damp membrane in a hybridization bag and incubate for 10 min at 37°C to enhance the luminescent reaction.
- 11. Expose for 2 hours at room temperature to X-ray film. Multiple exposures can be taken. Luminescence continues for at least 24 hours and signal intensity increases during the first hours.

Example 3: Transformation of Carrot Cells

Transformation of plant cells can be accomplished by a number of methods, as described above. Similarly, a number of plant genera can be regenerated from tissue culture following transformation. Transformation and regeneration of carrot cells as described herein is illustrative.

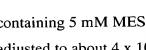
Single cell suspension cultures of carrot (*Daucus carota*) cells are established from hypocotyls of cultivar Early Nantes in B_5 growth medium (O.L. Gamborg et al., *Plant Physiol.* 45:372 (1970)) plus 2,4-D and 15 mM CaCl₂ (B_5 -44 medium) by methods known in the art. The suspension cultures are subcultured by adding 10 ml of the suspension culture to 40 ml of B_5 -44 medium in 250 ml flasks every 7 days and are maintained in a shaker at 150 rpm at 27 °C in the dark.

The suspension culture cells are transformed with exogenous DNA as described by Z. Chen et al. *Plant Mol. Bio.* 36:163 (1998). Briefly, 4-days post-subculture cells are incubated with cell wall digestion solution containing 0.4 M sorbitol, 2% driselase, 5mM MES (2-[N-Morpholino] ethanesulfonic acid) pH 5.0 for 5 hours. The digested cells are pelleted gently at 60 xg for 5 min. and washed twice in W5 solution containing 154 mM NaCl, 5 mM KCl, 125 mM CaCl₂ and 5mM glucose, pH 6.0. The protoplasts are suspended in MC solution

10

15

20



containing 5 mM MES, 20 mM CaCl₂, 0.5 M mannitol, pH 5.7 and the protoplast density is adjusted to about 4×10^6 protoplasts per ml.

15-60 µg of plasmid DNA is mixed with 0.9 ml of protoplasts. The resulting suspension is mixed with 40% polyethylene glycol (MW 8000, PEG 8000), by gentle inversion a few times at room temperature for 5 to 25 min. Protoplast culture medium known in the art is added into the PEG-DNA-protoplast mixture. Protoplasts are incubated in the culture medium for 24 hour to 5 days and cell extracts can be used for assay of transient expression of the introduced gene. Alternatively, transformed cells can be used to produce transgenic callus, which in turn can be used to produce transgenic plants, by methods known in the art. See, for example, Nomura and Komamine, Plt. Phys. 79:988-991 (1985), Identification and Isolation of Single Cells that Produce Somatic Embryos in Carrot Suspension Cultures.

An additional deposit of an E. coli Library, E. coli LibA021800, was made at the American Type Culture Collection in Manassas, Virginia, USA on February 22, 2000 to meet the requirements of Budapest Treaty for the international recognition of the deposit of microorganisms.

The invention being thus described, it will be apparent to one of ordinary skill in the art that various modifications of the materials and methods for practicing the invention can be made. Such modifications are to be considered within the scope of the invention as defined by the following claims.

Each of the references from the patent and periodical literature cited herein is hereby expressly incorporated in its entirety by such citation.